Economics and Philosophy, **33** (2017) 215–257 © Cambridge University Press
doi:10.1017/S0266267116000225 First published online 19 December 2016
cambridge.org/eap

THE PRIORITY VIEW

DAVID MCCARTHY*

Abstract: According to the priority view, or prioritarianism, it matters more to benefit people the worse off they are. But how exactly should the priority view be defined? This article argues for a highly general characterization which essentially involves risk, but makes no use of evaluative measurements or the expected utility axioms. A representation theorem is provided, and when further assumptions are added, common accounts of the priority view are recovered. A defence of the key idea behind the priority view, the priority principle, is provided. But it is argued that the priority view fails on both ethical and conceptual grounds.

Keywords: Priority view, prioritarianism, egalitarianism, utilitarianism, distribution

1. INTRODUCTION

According to the priority view, it somehow matters more to benefit people the worse off they are. Much has been written about how plausible the priority view is, and in particular, whether it is more plausible than egalitarianism. This article is mostly concerned with a different topic: what exactly is the priority view?

There have been two main ways of trying to understand the priority view. Early discussions mostly did not mention risk, and assumed that the priority view could be understood in risk-free terms. These discussions took it for granted that it makes sense to talk quantitatively about how well off people are. More recent discussions do discuss risk, and sometimes define the priority view in terms of risk. But they continue to talk quantitatively about how well off people are, and make heavy use of the expected utility axioms.

After some preliminary matters in sections 2 and 3, sections 4 and 5 explain why this article rejects these approaches to understanding the

* Department of Philosophy, University of Hong Kong, Pokfulam, Hong Kong. Email: davidmccarthy1@gmail.com URL: http://philosophy.hku.hk/?n=Main.StaffDM

priority view. The positive account of what the priority view amounts to is developed in sections 6 to 8. The result is a very general definition of the priority view which avoids quantitative talk about evaluative properties, is essentially about risk, but nevertheless makes no use of the expected utility axioms. Section 9 provides a representation theorem characterizing the priority view, so defined. This is the only technical section of the article, but the results will be summarized informally. Section 10 discusses how the priority view fits into our taxonomy of distributive theories. The main goal of the article is to try to understand the priority view, but section 11 ends by trying to assess its truth, and concludes that it is false.

The article makes significant use of views about egalitarianism and utilitarianism developed in McCarthy (2015) and McCarthy *et al.* (2016) respectively. Those works give accounts of egalitarianism and utilitarianism which are essentially about risk but make no use of quantitative evaluative measures or the expected utility axioms. The present article is therefore part of a more general approach to understanding the ethics of distribution.

2. TERMINOLOGY

Say that any two individuals who are equally well off have the same *welfare level*. We assume as given a set of welfare levels which we think of as the set of all possible welfare levels. The discussion is not tied to any particular interpretation of welfare, and we use 'welfare' as a cover-all term which can be interpreted in terms of happiness, preference satisfaction, objective goods, capability sets and so on. We will be exclusively interested in views which are somehow only concerned with the distribution of welfare levels, and not with such things as responsibility or desert.

We will be interested in risk. For simplicity we will only consider probability functions in which only finitely many outcomes are possible. A *prospect* is a probability function over welfare levels. Again for simplicity, we will take the background set of prospects to be the set of all such probability functions. With little change, the axioms, arguments and results of this article could be stated in a framework which uses general probability measures.[1] But this would greatly increase the technical difficulty.

We assume throughout a fixed, finite population of individuals $\mathbb{I} = \{1, \ldots, n\}$.[2] A *history* is an assignment of a welfare level to each member of $\mathbb{I}$. Thus a history h can be written in the form $h = [x_1, \ldots, x_n]$ where x_i

[1] Such a framework is provided in McCarthy *et al.* (2016). The present framework is essentially a simplification of this more general one.

[2] In section 9 we will, however, consider the relationship between social preorders relative to different populations. The account of variable population problems in McCarthy

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

is the welfare level of the ith individual. Again, this is not necessary, but we will assume that the set of histories is the set of all possible profiles of welfare assignments. A *lottery* is a probability function over histories, and the background set of lotteries is the set of all such probability functions.

I now need to explain the idea of a permutation acting on a lottery. The basic idea is that a permutation just swaps the identities of the individuals around. For example, if L is a lottery in which Ann gets a for certain and Bob has a one in two chance of b or c, then the lottery L' in which Ann gets a one in two chance of b or c and Bob gets a for certain is obtained by permuting Ann and Bob.

In more detail, a permutation of the population is a function which maps each individual onto some individual, with no two individuals mapped onto the same individual. Thus in jargon, a permutation is a bijection from the population to itself. To illustrate, let the population be $\mathbb{I} = \{1, 2, 3\}$. Then the map σ which maps 1 to 2, 2 to 3, and 3 to 1, that is, $\sigma 1 = 2$, $\sigma 2 = 3$ and $\sigma 3 = 1$, is a permutation. Its inverse σ^{-1} is just the inverse of the map. Thus $\sigma^{-1}1 = 3$, $\sigma^{-1}2 = 1$ and $\sigma^{-1}3 = 2$. We next need to define a permutation acting on a history. Given a history $h = [x_1, \ldots, x_n]$ and a permutation σ of $\mathbb{I}$, σh is the history defined by $\sigma h := [x_{\sigma^{-1}1}, \ldots, x_{\sigma^{-1}n}]$. In other words, in σh, i has the welfare level which $\sigma^{-1}i$ has in h. Thus in the example, if $h = [a, b, c]$, then $\sigma h = [c, a, b]$. Finally, given a lottery $L = [p_1, h_1; p_2, h_2; \ldots; p_m, h_m]$, where the h_i's are the various possible histories which could result and p_i is the probability of h_i, we define a lottery σL by $\sigma L := [p_1, \sigma h_1; p_2, \sigma h_2; \ldots; p_m, \sigma h_m]$. Roughly speaking, σL is the same lottery as L, except that the prospect faced by i under σL is the prospect faced by $\sigma^{-1}i$ under L. In the example, let $L = [\frac{1}{2}, [a, b, c]; \frac{1}{2}, [x, y, z]]$. Then $\sigma L = [\frac{1}{2}, [c, a, b]; \frac{1}{2}, [z, x, y]]$. This construction will later allow us to express ideas to do with anonymity.

We will later be considering populations of size one, and there it will be important to distinguish lotteries from prospects. Notationally, $p = [\frac{1}{2}, a; \frac{1}{2}, c]$ is the prospect in which welfare levels a and c are equally likely, whereas $L = [\frac{1}{2}, [a]; \frac{1}{2}, [c]]$ is the lottery in which the two histories $[a]$ and $[c]$ are equally likely, where in those histories the sole person has welfare levels a and c respectively. Of course, p is the prospect the sole individual faces under L, but p and L are not identical. We will often identify welfare levels and histories with degenerate prospects and lotteries. For example, we will often treat the welfare level b as the prospect in which welfare level b is certain, and the history $[b]$ as the lottery in which the history $[b]$ is certain.

et al. (2016) can straightforwardly be used to generate a prioritarian account of variable population problems given the account of prioritarianism to be developed in the sequel, but we will not pursue this.

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

The *better prospect relation* holds between two prospects p and p' just in case p is at least as good a prospect as p'. This relation can be thought of as encoding both intrapersonal and interpersonal comparisons. To illustrate, for any individual i, let $\mathcal{P}_i(L)$ be the prospect i faces under L. Then for any individuals i and j, identical or not, we have: $\mathcal{P}_i(L)$ is at least as good a prospect as $\mathcal{P}_j(L')$ if and only if L is at least as good for i as L' is for j. As this suggests, personal identity is taken to be irrelevant to comparisons between the prospects individuals face. If i and j face identical prospects, they are equally well off, even if they have different attitudes towards risk: any anxiety or excitement, for example, is already accounted for in their prospective welfare levels. I will sometimes refer to the relation 'at least as good for i as' which holds between lotteries as i's *individual betterness relation*. But to repeat, this just captures intrapersonal comparisons for i, and is encoded in the better prospect relation.

I will assume that the better prospect relation is a preorder on the set of prospects. A preorder is a binary relation which is reflexive and transitive. This allows the better prospect relation to be incomplete: thus there can be prospects p and p' such that it is not the case that p is at least as good a prospect as p', and also not the case that p' is at least as good a prospect as p. The point of allowing for incompleteness is to accommodate the common view that there may be significant limitations to such things as intrapersonal and interpersonal comparisons. I do not claim that this is the only or the best way to make room for that view, but it is the simplest and most natural. To mark the fact that this assumption is always in play, I will typically refer to the better prospect relation as the *individual preorder*.

To summarize, we are assuming that we are addressing a class of problems in which we are exclusively concerned with the distribution of welfare. Different theorists have different views about what welfare amounts to (e.g. happiness, achievement etc.), but we are assuming that all theorists agree that there are welfare levels (e.g. levels of happiness, levels of achievement etc.). But risk is endemic, so we have to consider not just situations in which individuals have particular welfare levels, but situations in which individuals face particular prospects, where it is a matter of risk which welfare levels they turn out to have. We are assuming that it makes sense to say that one prospect is at least as good a prospect as another (e.g. a half chance of happiness and a half chance of great happiness is at least as good as misery for certain), and we are making the very weak assumption that the corresponding relation is a preorder. Thus we are allowing that there could be extreme, mild or no incomparability between different welfare levels or prospects. We are also neither accepting nor rejecting the assumption that the better prospect relation, or individual preorder, satisfies any further axioms to do with risk, such as the expected utility axioms. Our framework is therefore extremely general.

The *betterness relation* holds between two lotteries L and L' just in case L is at least as good as L'. I will assume that the betterness relation is a preorder on the set of lotteries, and will therefore typically refer to it as the *social preorder*.

I will sometimes write the social preorder as $\succsim$, so $L \succsim L'$ means that L is at least as good as L'; $L \succ L'$ means that L is better than L', and $L \sim L'$ means that they are equally good. Similarly, I will sometimes use $\succsim_{\mathbb{P}}$ (so labelled because it preorders prospects) to denote the individual preorder, with similar notational variants.

3. THE PRIORITARIAN PLATITUDES

Parfit (1991) brought the priority view into prominence among philosophers. Similar views had been discussed before,[3] but it was Parfit who introduced the term 'the priority view' (as well as its synonym 'prioritarianism'), so any definition of it has to largely defer to what he said about it. But Parfit did not come anywhere near providing a full, explicit definition. Rather, he said a number of things about the priority view which together should be seen as an attempted implicit definition. For example, he famously said that according to the priority view, benefiting the worse off matters more (Parfit, 1991, 19). It is not the case that he has provided a full definition of the priority view, and is here pointing out something which the priority view logically implies. Rather, his claim is supposed to get us to understand what the priority view amounts to in the first place. Thus I will refer to such claims as 'according to the priority view, benefiting the worse off matters more' as *prioritarian platitudes*, and think of them as collectively defining the priority view, if anything does. I call them 'platitudes' to indicate that while there are plenty of ways we may disagree with Parfit, thinking that the prioritarian platitudes systematically misdescribe the priority view is not one of them.

What I take to be a more or less exhaustive list of prioritarian platitudes is provided in McCarthy (2013), but here I will just summarize the prioritarian platitudes as we go along. But the main idea is that the priority view is that theory, which when interpreted as the priority view, vindicates the prioritarian platitudes both well enough and better than any other theory. This is nothing more than an informal application of the so-called Ramsey–Carnap–Lewis treatment of implicit definition (see e.g. Lewis, 1970, Jackson, 2000).

Parfit's initial discussion of the priority view assumed that we could make sense of the priority view by talking quantitatively about various evaluative properties, such as goodness for people or goodness, and

[3] See e.g. Broome (1989) and Hurley (1989). In welfare economics, the position arguably goes at least as far back as Atkinson and Stiglitz (1980, 340).

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

without talking about risk. Much of the early work trying to elaborate on the priority view did the same.[4] But McCarthy (2013) argues that this approach to understanding the priority view is bound to be unsuccessful. Here I will take this conclusion for granted and will not repeat the arguments for it. Instead, I propose to try to understand the priority view by talking about risk and without talking quantitatively about evaluative properties.

However, more recent literature has tried to understand the priority view by augmenting quantitative talk about evaluative properties with talk about risk. My proposal will therefore seem perverse without a brief sketch of the difficulties this approach faces. Rather than look at all of the ways quantitative evaluative measures have been used in conjunction with risk to try to understand the priority view, we will focus on the main prioritarian idea.

4. WELFARE UNITS

The basic idea of the priority view is that it is more urgent to benefit people the worse off they are, adding that this has nothing to do with inequality. The first part of this clause has often been interpreted as what is often known as the Pigou–Dalton principle.

(PD) It is better to give a unit of welfare to the worse off of two people.

This interpretation presupposes that units of welfare are well-defined; or as I will say, that welfare is unit-measurable. For this to hold, it is sufficient for welfare measures to be unique up to positive affine transformation, but I will ignore this, and pretend they are unique. Section 8.2 will say more about how (PD) has typically been used to develop a full definition of the priority view, but for now we will focus on (PD) itself.

Perhaps Parfit's most fundamental claim about the priority view was that it is a novel distributive theory and expresses an ethically plausible idea. Any interpretation of the priority view should try hard to vindicate this claim. Now (PD) is a fine-sounding idea, but does it really express an ethically plausible idea?

There are two ways in which we could become convinced that it does. Perhaps the plausibility of (PD) is simply transparent to intuition. Alternatively, some argument could be given for (PD). We will look at both possibilities, but first some background.

When Parfit first discussed the priority view, he did not mention risk and just assumed that welfare is unit-measurable. He was thus tacitly adopting the well-known position that what fixes the units has nothing to

[4] See e.g. Jensen (2003) and Tungodden (2003), although these writers were certainly aware of the complexities introduced by risk.

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

do with risk, so that it is a genuine question how it interacts with risk. For example, if one prospect has a higher expectation of welfare than another, it is an open question whether it is a better prospect. Call the welfare-measure this view has in mind 'the natural measure'.

In his most recent discussion, however, Parfit (2012) essentially assumes what Broome (1991) calls *Bernoulli's hypothesis*. This presupposes that the individual preorder satisfies the expected utility axioms, so that it can be represented by a vNM utility function, and then claims that the vNM utility function measures welfare. This implies that welfare is unit-measurable. But what fixes the units now involves risk, and the interaction between risk and welfare is given by definition: one prospect is better than another if it yields a higher expectation of welfare. Refer to the corresponding welfare-measure as 'the vNM measure'.[5]

Parfit's slide from the natural measure to the vNM measure is matched by the literature. The early literature on the priority view mostly made no mention of risk, and was therefore tacitly adopting the natural measure. More recent literature has adopted the vNM measure largely without comment.

Given this state of affairs, one would expect the two measures to be equivalent. But it is hard to see any guarantee of this.[6] Consider natural and vNM measures of preference satisfaction. The former has most often been seen as constructed from strength of preference judgements, such as the judgement that the preference for \$1 over \$0 is stronger than the preference for \$11 over \$10.[7] But there is empirical evidence that when so understood, the natural and vNM measures are not equivalent (Bouyssou and Vansnick 1988; Pennings and Smidts 2000).

Now many philosophers do not understand welfare in terms of preference satisfaction. But they may still form judgements about when one welfare difference is greater than another, and when one gamble over welfare levels is better than another. But as in the preference case, it is hard

[5] When I talk about *the* natural and vNM measures, I am harmlessly suppressing the fact that they will only be unique up to positive affine transformation. In addition, there is more than one way of thinking about the natural measure, and the difference-based measure to be discussed below may be one of a plurality of measures. See e.g. Abdellaouia *et al.* (2007) for further discussion. But the existence of a plurality of natural measures would only make the difficulties I am about to discuss for the motivation of (PD) more severe, so I will suppress this issue as well.

[6] This is essentially the basis of the well-known Sen–Weymark objection to Harsanyi's (1955, 1977) claim that his famous aggregation theorem should be given a utilitarian interpretation. Harsanyi was assuming or arguing for the vNM measure as the correct welfare measure. But Sen (1976, 1977) and Weymark (1991) objected that utilitarianism is traditionally defined in terms of the natural measure, and complained that Harsanyi had failed to show that the two measures are equivalent.

[7] For an axiomatization of measures resulting from so-called difference orderings, see e.g. Suppes and Winet (1955).

to see a guarantee that the resulting natural and vNM measures would be equivalent.[8]

But this means that it is a serious possibility that (a) (PD) understood with respect to one measure is not equivalent to (PD) understood with respect to the other measure; and, as will be explained in more detail in section 8.2, (b) the same social ordering is utilitarian according to one measure, and prioritarian according to the other. We now consider four rationales for (PD).

First, the most common view among prioritarians is that (PD) is transparently plausible. But (PD) is seriously ambiguous. Indeed, several writers have claimed that it is vague or indeterminate what the true welfare measure is, and this vagueness or indeterminacy will transmit to (PD). But even if some subtle consideration were to rule in favour of a particular welfare measure, it is hard to see how this would help the case for the transparent plausibility of (PD). For given that (a) and (b) are serious possibilities, and given that writers such as Parfit appear to be insensitive to this fact, it is hard to believe that their intuitions in favour of (PD) are reliable. I am, of course, not asserting that (PD) is implausible. Rather, in the absence of a convincing backstory about why there is a sufficiently determinate welfare measure, and an explanation of why intuition about the distribution of welfare units is reliable, I simply claim that truth of (PD) is deeply unclear. For example, Broome (1991)'s treatment of what makes something a welfare measure, elaborated on by Jensen (1995), is perhaps the most original and detailed treatment of what makes something a welfare measure, and he concluded that intuition notwithstanding, in the context of an otherwise prioritarian framework, (PD) is false more or less as a matter of meaning. Section 11 will discuss Broome's argument in more detail. But for now, regardless of whether Broome's argument stands up to scrutiny, the idea that one can simply intuit the plausibility of (PD) is hard to take seriously.

Second, define (PD_{nat}) and (PD_{vNM}) in the same way as (PD), but understood with respect to natural and vNM units respectively. Someone might judge that (PD_{vNM}) is intuitively plausible, accept an argument for the vNM measure as the true welfare measure, or follow those who suggest adopting it as a convention in light of vagueness or indeterminacy about what the true measure is, and conclude with (PD). This person trusts her intuitions about the distribution of vNM units, and unlike the

[8] As far as I am aware, the only acknowledgement of this possibility in the literature on the priority view is in the groundbreaking work of Rabinowicz (2001, 2002). But Rabinowicz rejects the idea that both measures could exist on the grounds that the information needed to obtain both is too demanding, and settles for using the vNM measure. However, I think he has imposed overly demanding constraints on the natural measure. This measure emerges merely from an ordering of welfare differences which obeys some natural axioms and does not, contra Rabinowicz, have to say anything about risk.

previous position, intuition about the distribution of welfare units plays no real role. In response, I find it most unclear why intuition rallies behind (PD_{vNM}) rather than, say, (PD_{nat}), but more generally, the whole approach is surely misguided. The existence of the vNM measure follows from some quite subtle axioms and a fairly demanding mathematical construction. The measure interacts neatly with risk just because it has been designed to, but there is far too much mathematical complexity behind it for it to be the vehicle of reliable intuitions about distribution.

Third, instead of claiming that (PD) has intuitive support, one might offer an argument for it. Thus Rabinowicz (2001, 2002) accepts the vNM measure, then argues that (PD) is motivated by Rawls's separateness of persons objection to utilitarianism (see also Adler 2011, 314 ff.). However, Rawls was actually criticizing the impartial spectator argument. For Rabinowicz's rationale to succeed, the implicit welfare measure in the impartial spectator argument must be equivalent to the vNM measure. This is not obvious, but let that pass and assume equivalence. The problem now is that arguments for utilitarianism based on variants of Harsanyi's aggregation theorem are much more powerful than the impartial spectator argument. But they assume nothing like the construction Rawls was criticizing, so plainly no rationale for (PD) has been provided.

Fourth, perhaps some more complicated argument could be offered for (PD_{vNM}), followed by the acceptance of the vNM metric, with (PD) in its wake. But we have been interested in using (PD) to interpret the basic idea behind the priority view. That precludes using complicated arguments for it, otherwise (PD) will count as derived rather than basic.

In summary, combining talk about welfare units with talk about risk seems to create – or at least reveal – more problems than it solves. It shows there is an ambiguity in (PD) which undermines the claim that (PD) is transparently plausible. Adopting the vNM measure at least resolves the ambiguity, and at least provides a measure which interacts conveniently with risk. But it does not help make the case for the transparent plausibility of (PD). Thus starting with (PD) is not a promising way of trying to interpret the priority view, and I will not be trying to make sense of the priority view by talking about welfare units, or any other evaluative measure.

5. STRATEGY

Instead, I will try to offer an axiomatization of the priority view within the framework of section 2. By contrast, most of the literature on the priority view has been informal. The priority view has often been explained metaphorically; illustrated by an apparently unending series of vignettes; defined by a sometimes loose informal description of a mathematical

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

function; or defined directly by an explicit formula. But particularly when risk is involved, this situation is hard to understand.

Take *one-person cases*, lotteries where the population is fixed and of size one. An immense amount of progress has been made for this situation by the development of utility theory and decision theory, and the method is now standard: relatively simple axioms, typically involving comparatives, are motivated, then a representation theorem is proved. The point of a representation theorem is that it provides a convenient summary of what the axioms jointly imply, which otherwise would typically be very difficult to understand. Now contrast the situation in which someone were to present a new decision theory by simply presenting a formula which defines it, on the grounds that this formula implied one or two ideas that the person found appealing. The problem with this is that one has little idea what the formula is really saying; in particular, it is far from obvious what, if any, set of simple and motivated axioms is sufficient for the formula to hold. These remarks are trivial, but if one-person cases involving risk demand this kind of approach, many-person cases plainly do as well.

A further difficulty with using the vNM metric to define the priority view is that it presupposes that the individual preorder satisfies the expected utility axioms. But just about all of the expected utility axioms have been disputed, and the basic prioritarian idea does not sound as if it depends on the expected utility axioms. It would be surprising, to say the least, if accepting the basic prioritarian idea somehow precluded one from holding that the individual preorder satisfies some nonexpected utility theory for example. It is therefore natural to explore whether one can make sense of the priority view without using the expected utility axioms, but working with the vNM measure precludes this from the outset. Thus as well as making no use of quantitative evaluative measurements, I will be making no use of the expected utility axioms.[9]

The next three sections are organized roughly along the lines of asking how the priority view should be understood as contrasting with the three main distributive theories on the market, as well as discussing related prioritarian platitudes. The three theories are *ex post* egalitarianism (which values equality of outcome); *ex ante* egalitarianism (which values equality of prospects); and utilitarianism. The first two contrasts will be brief as the topics are familiar; it is in the contrast with utilitarianism that the distinctive and puzzling feature of the priority view emerges,

[9] I am, of course, assuming that the betterness relation and the better prospect relation are preorders. For convenience, by the expected utility axioms I will mean the further axioms needed to guarantee an expected utility representation. In the simple framework adopted here, these are the completeness axiom, the Archimedean axiom, and one of the standard independence axioms, to be defined below. Nonexpected utility theory is typically characterized by the rejection of the independence axiom.

and in its wake, a much neglected topic in the ethics of distribution. Although it is possible to combine different distributive concerns, I will reserve 'the priority view' for the pure theory which is indifferent to equality of outcome, indifferent to equality of prospects, but somehow anti-utilitarian.

To develop this account I will be drawing heavily on McCarthy (2015) and McCarthy *et al.* (2016). These offer accounts of egalitarianism and utilitarianism respectively, with no use of quantitative evaluative measurements and minimal use of expected utility theory. The account of the priority view which follows is therefore very much part of a package deal. But because this work is recent, section 9 will illustrate how this account of the priority view works in the more familiar expected utility environment.

6. ANTERIORITY

In this section, by 'egalitarianism' I will mean the view which values equality of outcome. One of Parfit's claims about the priority view is as follows. Page references are to Parfit (1991).

> (E) There is an ethically fundamental contrast between the priority view and egalitarianism (23).

This is made more concrete by further claims.

> (E_1) According to the priority view, and in contrast with egalitarianism, it does not matter how well off people are relative to one another (22–3).
>
> (E_2) According to the priority view, and in contrast with egalitarianism, benefits have absolute value (23).
>
> (E_3) Egalitarianism, unlike the priority view, is subject to a levelling-down objection (17).

Interpreting the priority view (and egalitarianism) in a way which makes sense of (E_1) through (E_3) will go far towards vindicating (E). But (E_1) through (E_3) themselves need interpretation, for some of the concepts they contain are unclear. We consider them in order.

6.1. Relativity

To make sense of (E_1), we need to interpret the idea that it does not matter how well off people are relative to one another. A clue is provided by an example due to Myerson (1981).[10] Assume a population of two people, *A*

[10] For discussion of Myerson's example, see among others Meyer and Mookherjee (1987), Broome (1989, 1991, 2015), Ben-Porath *et al.* (1997); Rabinowicz (2002); Fleurbaey (2010, 2015) and McCarthy (2015).

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

and B, and let a and b be distinct welfare levels.

L_E	$\frac{1}{2}$	$\frac{1}{2}$
A	a	b
B	a	b

L_F	$\frac{1}{2}$	$\frac{1}{2}$
A	a	b
B	b	a

A and B each face the same prospect in both lotteries, namely $[\frac{1}{2}, a; \frac{1}{2}, b]$. But the prospects they face are combined in different ways in L_E and L_F. Those who are unconcerned with how well off people are relative to one another will naturally regard these different combinations as equally good. But those who are concerned may well think that one combination is better than the other. For example, egalitarians will think that L_E is better than L_F: each individual faces the same prospect, but L_E guarantees equality of outcome while L_F guarantees inequality. Those like Nietzsche who are concerned with relativities by thinking equality a bad thing will think L_F better than L_E.

One natural way of expressing the idea that it does not matter how well off people are relative to one another is to use the following concept, formally defined in McCarthy (2015) under the label 'weak separability across individual lotteries'. Suppose that two lotteries L and L' are identical if we ignore some individual i, and i nevertheless faces the same prospect in L as in L'. Then the betterness relation satisfies *weak prospect separability* if in all such cases, L and L' are equally good. For example, L_E and L_F are identical if we ignore B, and B faces the same prospect in each. So weak prospect separability implies that L_E and L_F are equally good. Thus the natural idea is to at least partially interpret the idea that it does not matter how well off people are relative to one another by claiming that it implies weak prospect separability.

However, I now prefer to use a variation on the same theme, borrowed from McCarthy *et al.* (2016).

> **Anteriority** For all lotteries L and L': if $\mathcal{P}_i(L) = \mathcal{P}_i(L')$ for all $i \in \mathbb{I}$, then $L \sim L'$.

In words: if every individual faces the same prospect in L as in L', then L and L' are equally good.

Anteriority and weak prospect separability are closely related: Anteriority obviously implies weak prospect separability, while weak prospect separability plus transitivity of 'as good as' implies Anteriority. So there is not much to choose between them. Nevertheless, Anteriority is simpler to understand, and I suggest we accept

> (A) *It does not matter how well off people are relative to one another* $\Rightarrow$ the social preorder satisfies Anteriority.

This offers an interpretation of the slightly vague and informal idea conveyed by the italicized expression, and so cannot be proved.

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

Nevertheless, it is hard to deny that anyone who accepts or claims to understand the informal idea should also accept Anteriority.

Now notice that (A) only offers a partial interpretation. I am not making the stronger claim that the informal idea is equivalent to Anteriority. I am not denying it either, but whether we should make the stronger claim turns out to be a conceptually and technically complicated matter. However, as a practical issue this is not too important. The only live view according to which it does matter how well off people are relative to one another is the idea that equality is valuable. But anyone who thinks that will have to think that in Myerson's example, L_E is better than L_F, violating Anteriority. Any view according to which some relativities do matter which somehow does not result in a failure of Anteriority is therefore going to have to be strange and unfamiliar, so I will set this topic aside.

Putting all this together, we can vindicate (E_1) by accepting

> (B) One fundamental contrast between the priority view and egalitarianism is that the priority view accepts Anteriority while egalitarianism rejects it.[11]

Several writers have concluded that it is not possible to distinguish the priority view from egalitarianism in the following sense: the set of prioritarian betterness relations is included in the set of egalitarian betterness relations; there is no case about which the two theories necessarily disagree.[12] But we should try to avoid this conclusion given the implicit definition approach to understanding the priority view sketched in section 3. For that method puts us under interpretative pressure to make sense of as many prioritarian platitudes as possible. Since so much of what Parfit said about the priority view rested on contrasting it with egalitarianism, (B) enjoys a considerable advantage. For it is based on a simple and clear case which egalitarian and prioritarian betterness relations necessarily disagree about, namely Myerson's example.

In addition, (B) goes some way towards vindicating (E), since Anteriority is clearly a natural and ethically fundamental principle. We will add further support for (E) in section 6.3 when we will discuss the levelling down objection.

6.2. The absolute value of benefits

Consider now (E_2). To use this platitude, we have to interpret the idea that benefits have absolute value, and it is not so clear how to do that. The

[11] I am not aware of anyone analysing the contrast between the priority view and egalitarianism in terms of Anteriority. But for varying degrees of sympathy towards ideas in the ballpark of the present proposal, see e.g. Broome (1989, 1991, 2015), Fleurbaey (2015) and Rabinowicz (2001, 2002).

[12] See e.g. Tungodden (2003). However, this article did not consider risk.

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

literature has generally tried to do so by taking it for granted that it makes sense to speak quantitatively about goodness. The rough idea is then that benefits have absolute value just in case providing someone with a particular benefit produces the same amount of goodness, independently of the context. However, as already mentioned I am rejecting this general approach to understanding the priority view, and for detailed criticisms of this idea in particular, as well as references to those who have adopted it, see McCarthy (2013). Here instead is a different interpretation.

Assume a population of two people, A and B. Assume a fair coin, and let a, b, x and y be welfare levels, with b higher than a. Consider a lottery $L = [\frac{1}{2}, [a, x]; \frac{1}{2}, [a, y]]$. Suppose we can now give b to A, rather than a, in either state of nature. If we give it to A in the first, the context will be that B has x. If we give it to A in the second, the context will be that B has y. No matter how we understand benefits, the benefit to A will be the same in each case: b rather than a. So the claim that benefits have absolute value – a value independent of context – is naturally exemplified by the claim that the two options are equally good. Thus it is natural to interpret the idea that benefits have absolute value as exemplified by the claim that the two lotteries $L' = [\frac{1}{2}, [b, x]; \frac{1}{2}, [a, y]]$ and $L'' = [\frac{1}{2}, [a, x]; \frac{1}{2}, [b, y]]$ are equally good.

We need to generalize, but how to do this is obvious.

(C) *Benefits have absolute value* $\iff$ the social preorder satisfies weak prospect separability.

Again, this offers a formal interpretation of the somewhat vague italicized expression, and does not assume anything about goodness measures. Since Anteriority is equivalent to weak prospect separability given that the social preorder is transitive, (B) also vindicates (E_2).

6.3. Levelling down

Parfit's so-called levelling-down objection to egalitarianism claims that according to egalitarianism, things are in one way better if everyone is brought down to the level of the worst off. He suggested that this conclusion is absurd, and the fact that it avoided the levelling-down objection made a large part of his case for the priority view.

The levelling-down objection has had a mixed reception; see e.g. Holtug (2010, ch. 7) for further discussion. I will mention three criticisms. (a) Some have claimed that some forms of egalitarianism are immune to it since they entail that it is never on balance better to level down (see e.g. Fleurbaey 2015). (b) Some have claimed that the levelling-down objection relies on an obscure notion of being 'better in some respect' (e.g. Brown 2003). (c) Others have claimed that egalitarianism can be defined in a way which is not subject to the objection (e.g. Jensen 2003).

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

However, McCarthy (2015) offers a version of the levelling-down objection and claims that it is a decisive objection to egalitarianism. The basic idea is that with a mild continuity condition, because they think L_E is better than L_F, egalitarians are committed to the existence of some lottery M which is (i) like L_E, in particular still guaranteeing equality, except that every possible welfare level is a little worse, and (ii) still better than L_F. Thus the prospect each individual faces under M is strongly dominated by the prospect the individual faces under L_F. McCarthy (2015) argues that this egalitarian commitment is indefensible. Thus unlike response (a) to the levelling-down objection, the objection involves an on-balance betterness judgement; unlike response (b), the objection does not use the notion of being better in some respect; and unlike response (c), it relies on a very weak claim about egalitarianism, namely that egalitarians will think L_E is better than L_F, a claim which has been endorsed by everyone I know of who has considered the example.

I do not of course expect egalitarians to immediately concede. But it is a serious objection, and one which to my mind is more effective than Parfit's original objection. But the present point is that it arises because egalitarians reject Anteriority. A quick glance at the argument shows that it will apply to any view which rejects Anteriority in thinking that one lottery can be better than another despite the fact that Anteriority says they are equally good. Conversely, any view which accepts Anteriority is invulnerable to at least this route to the objection. Thus a bonus of accepting (B) is that it brings out an asymmetry between egalitarianism and prioritarianism with respect to the levelling-down objection, and thus at least partially vindicates (E_3). However, we will revisit the levelling-down objection in section 11.1, where it will turn out that although prioritarianism most likely avoids the letter of the levelling-down objection, it does not avoid the spirit of it.

In short, accepting (B) leads to the vindication of (E_1) and (E_2), and at least the partial vindication of (E_3). Overall, it therefore does a good job at vindicating (E), for recall that vindication does not have to be perfect. Thus ignoring the fact that there are minor variants of it, I suggest that Anteriority is one of the core components of the priority view.

7. TWO-STAGE ANONYMITY

As well as the view just discussed, which values equality of outcome, there is also the view which values equality of prospects. The essence of this view is encapsulated in the following example, due to Diamond (1967), which compares the lottery L_F introduced in section 6.1 with a

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

new lottery, L_U. We assume a is a higher welfare level than b.

L_F	$\frac{1}{2}$	$\frac{1}{2}$
A	a	b
B	b	a

L_U	$\frac{1}{2}$	$\frac{1}{2}$
A	a	a
B	b	b

Assuming that the identities of individuals are irrelevant, $[a,b]$ is exactly as good as $[b,a]$, so the two lotteries produce equally good outcomes. But there is equality of prospects in L_F and inequality of prospects in L_U. For this reason, some regard L_F as better than L_U.[13] This idea is the hallmark of what is sometimes known as *ex ante* egalitarianism.

We are trying to understand the priority view as a pure theory, not a hybrid which incorporates already familiar anti-utilitarian ideas. In addition, nothing like the idea that L_F is better than L_U appeared in Parfit (1991), so it is natural to interpret the priority view as assuming that L_F and L_U are equally good.

Two principles are usually adopted to generalize the idea that L_F is exactly as good as L_U. The first formalizes the idea that the identities of individuals are irrelevant.

> **Anonymity** For all lotteries L and permutations σ on the population: $L \sim \sigma L$.

The second is some variant of the idea that the social preorder satisfies one of the independence axioms from expected utility. The best known is the following.

> **Strong Independence** For all lotteries L, L' and M, and all $\alpha \in (0,1)$: $L \succsim L' \iff \alpha L + (1-\alpha)M \succsim \alpha L' + (1-\alpha)M$

These are jointly sufficient to entail that $L_F \sim L_U$, but McCarthy *et al.* (2016) adopt a principle which is strictly weaker than the conjunction of Anonymity and Strong Independence, namely

> **Two-Stage Anonymity** For all lotteries L and M, all $\alpha \in [0,1]$, and all permutations σ: $\alpha L + (1-\alpha)M \sim \alpha(\sigma L) + (1-\alpha)M$

To illustrate, let L be the history (degenerate lottery) $[b,a]$, M be the history $[a,b]$, σ be the permutation that swaps A and B, and $\alpha = \frac{1}{2}$. Then $\alpha L + (1-\alpha)M = \frac{1}{2}[b,a] + \frac{1}{2}[a,b] = L_F$, and $\alpha(\sigma L) + (1-\alpha)M = \frac{1}{2}[a,b] + \frac{1}{2}[a,b] = L_U$. Hence Two-Stage Anonymity entails that L_F is exactly as good

[13] The literature on this topic is vast. Early key works in welfare economics and philosophy respectively are Diamond (1967) and Broome (1990). The idea that L_F is better than L_U is often taken to reflect the view that fairness matters in the distribution of goods; see e.g. Broome (1991). McCarthy (2015) argues that this is a mistaken way of representing the idea.

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

as L_U. But it does not require the social preorder to satisfy Strong Independence.

Roughly speaking, the appeal of Two-Stage Anonymity is that it enables one to capture the natural idea that if two lotteries guarantee the same 'anonymous outcome', where for example $[a,b]$ and $[b,a]$ represent the same anonymous outcome (the welfare profiles are identical, except that the individuals are swapped around), then the two lotteries are equally good. This idea is usually captured by assuming Strong Independence and Anonymity. But Two Stage-Anonymity is significantly weaker. In particular, it allows the social preorder to violate Strong Independence, thus making room for a much wider range of views about the social preorder (and consequently, we will see later, the individual preorder). Thus Two-Stage Anonymity allows (but does not require) the individual and social preorders to satisfy various nonexpected utility theories. The significance of this will come out in section 9.

In summary, I suggest that Two-Stage Anonymity is another one of the core principles of the priority view. So far, there is nothing very new, as McCarthy *et al*. (2016) argue that Anteriority and Two-Stage Anonymity are core principles of utilitarianism. We now turn to what I take to be the anti-utilitarian component of the priority view, which is what is distinctive about the priority view.

8. THE PRIORITY PRINCIPLE

In what follows, section 8.1 explains the content of the proposed anti-utilitarian principle, which I call the Priority Principle. Section 8.2 elaborates on the standard account of the priority view, and discusses how the Priority Principle relates to it. The remaining sections 8.3 to 8.6 try to offer a defence of the Priority Principle.

8.1. Content

Parfit says little about utilitarianism. But when he says that according to the priority view, benefiting the worse off matters more, he clearly intends to be setting up a contrast with utilitarianism: benefiting the worse off matters more than utilitarianism says it does.

Parfit (1991, 23) also emphasizes that the main idea behind the priority view is intrapersonal. Commentators such as Rabinowicz (2001, 2002) have interpreted this to mean that the priority view has distinctive implications in one-person cases. Since inequality is never an issue in one-person cases, the contrast between utilitarianism and the priority view should be at its clearest in such cases.

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

In one-person cases, I take utilitarianism to be equivalent to

(RP) In one-person cases, L is at least as good as L' if and only if the prospect the sole person faces in L is at least as good as the prospect she faces in L'.

That done, what natural principle could be thought of as being more concerned with the worse-off in one-person cases than utilitarianism? In risk-free cases, it is clear that there should be no disagreement between utilitarianism and the priority view in one-person cases: both accept that in one-person cases, a history (a degenerate lottery) $h = [a]$ is at least as good as a history $h' = [b]$ if and only if a is at least as good a welfare level (degenerate prospect) as b.

Matters are different, however, when it comes to risk. Let a, b and c be three welfare levels, ranked from highest to lowest. Consider the two lotteries $L = [\frac{1}{2}, [a]; \frac{1}{2}, [c]]$ and $h = [b]$, and suppose the two prospects b and $[\frac{1}{2}, a; \frac{1}{2}, c]$ are equally good prospects. It follows from (RP), and hence utilitarianism, that h and L are equally good. Suppose now we want to give relatively greater weight to lower welfare levels than utilitarianism does. Then it is natural to say that in comparison with utilitarianism, it becomes relatively less important to secure welfare level a and relatively more important to avoid welfare level c. The net result is that given that b and $[\frac{1}{2}, a; \frac{1}{2}, c]$ are equally good prospects, the history h is better than the lottery L. Thus we arrive at what I take to be the paradigm of the anti-utilitarian component of the priority view.

(PP) In one-person cases, for any welfare levels x, y and z: (i) $[x] \succsim [y]$ if and only if $x \succsim_{\mathbb{P}} y$; and (ii) if $x \succ_{\mathbb{P}} y \succ_{\mathbb{P}} z$ and $y \sim_{\mathbb{P}} [\frac{1}{2}, x; \frac{1}{2}, z]$, then $[y] \succ [\frac{1}{2}, [x]; \frac{1}{2}, [z]]$.

We would like to generalize both (RP) and (PP) to cases where the population is of any fixed size. The following idea is natural: if $h = [b]$ is better than $L = [\frac{1}{2}, [a]; \frac{1}{2}, [c]]$ in the one-person case, then on any sane view (utilitarian, prioritarian, egalitarian and so on), in the n person case, $h' = [b, \ldots, b]$ will be better than $L' = [\frac{1}{2}, [a, \ldots, a]; \frac{1}{2}, [c, \ldots, c]]$. We thus obtain generalizations of (RP) and (PP), where we now go back to assuming an arbitrary but fixed population $\mathbb{I}$. To state these, note the obvious fact that if L is a lottery which guarantees perfect equality, then each member of the population faces the same prospect under L. For example, under the lottery L' just defined, each person faces the prospect $[\frac{1}{2}, a; \frac{1}{2}, c]$.

Reduction to Prospects For any lotteries L and L' which each guarantee equality, $L \succsim L' \iff P_L \succsim_{\mathbb{P}} P_{L'}$, where P_L is the prospect each person faces under L, and $P_{L'}$ is the prospect each person faces under L'.

McCarthy *et al.* (2016) argue that Reduction to Prospects is the fundamental connection between individual and social welfare according to utilitarianism.

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

Priority Principle For any welfare levels x, y and z: (i) $[x, ..., x] \succsim [y, \ldots, y] \iff x \succsim_{\mathbb{P}} y$; and (ii) if $x \succ_{\mathbb{P}} y \succ_{\mathbb{P}} z$ and $y \sim_{\mathbb{P}} [\frac{1}{2}, x; \frac{1}{2}, z]$, then $[y, \ldots, y] \succ [\frac{1}{2}, [x, \ldots, x]; \frac{1}{2}, [z, \ldots, z]]$.

I suggest that the Priority Principle is the fundamental anti-utilitarian component of the priority view. The formulation of the Priority Principle is a somewhat cautious generalization of (PP). Section 9 will discuss some further ways (PP) might be generalized.

8.2. Anticipation

It will help to review what is new about the Priority Principle, and what has been anticipated. The following account of the priority view, in part or in whole, stated in lesser or greater degrees of formality, has become common.[14] When part (i) by itself, or parts (i) and (ii) together, are stated as a definition of the priority view, the definition is presupposing the existence of a welfare measure g defined on welfare levels. For any individual i, g_i is then the function defined on histories by $g_i(h) = g(x_i)$ for any history $h = [x_1, \ldots, x_n]$. Thus g_i measures i's welfare in a given history.

Standard definition of the priority view (i) For some increasing and strictly concave w, the best available history maximizes the value of $w \circ g_1 + \ldots + w \circ g_n$.[15] (ii) For some increasing and strictly concave w, the best available lottery maximizes the expected value of $w \circ g_1 + \ldots + w \circ g_n$. (iii) g is the vNM measure.

Three remarks about this definition are worth making. First, part (i) of the definition has sometimes been seen as a full definition of the priority view. But regardless, it is easy to see roughly how it captures some prioritarian ideas. For the additive form has been seen as reflecting the anti-egalitarian aspect of the priority view, and the presence of the strictly concave function w has been seen as reflecting (PD). Nevertheless, there is quite some distance between these rough ideas and a precise axiomatization of (i).

Second, some writers seem to think that (ii) follows from (i) simply by appealing to expected utility theory. Many people have said this, but I will mention only two influential instances. Otsuka and Voorhoeve (2009, 195) say that prioritarians assess the 'moral value' of outcomes by attaching 'positive but decreasing marginal importance to each person's utility' (this is essentially (i)). By following 'orthodox decision theory',

[14] I believe that Rabinowicz (2001, 2002) was the first to argue for the entirety of this as the way to define the priority view, and that McCarthy (2008) was the first to provide an axiomatization of it.

[15] $w \circ g_i$ is the function defined on histories first by applying g_i, then w. Thus $(w \circ g_i)(h) = w(g_i(h))$.

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

prioritarians then 'maximize the probability-weighted sum of the moral value of these outcomes' (as in (ii)). Similarly, Parfit (2012, 404) says that prioritarians aim to maximize 'the sum of weighted benefits' (as in (i)). He then says that we should maximize expected goodness (his term is 'expectable goodness') to arrive at the conclusion that prioritarians aim to maximize 'the sum of expectable weighted benefits' (as in (ii)). However, the difficulties with this appeal were noted in Broome (1989) and discussed at length by Rabinowicz (2001, 2002) and McCarthy (2008); see also McCarthy (2013). The basic point is that (i) does not provide one with a measure of the 'moral value' or 'goodness' of outcomes according to prioritarians; but even if it did, merely appealing to 'orthodox decision theory' would not tell you to maximize the expectation of that measure anyway, and thus it would not take you to (ii). To get from (i) to (ii) we need a separability axiom which has a prioritarian motivation, one version of which was provided in McCarthy (2008), in addition to expected utility theory.

Third, does (ii) of the standard definition imply that the priority view implies the Priority Principle? Not by itself. Suppose g is the natural measure, and suppose that the individual preorder satisfies the expected utility axioms, so that the vNM measure is well-defined. As noted in section 4, there is no guarantee that the natural and vNM measures coincide. For example, it might well turn out that the individual preorder encodes risk-aversion with respect to the natural measure. In other words, it might turn out that getting some welfare level for certain is a better prospect than facing any genuine gamble over welfare levels which provides an equal expectation of welfare, as measured by the natural measure. To the extent that any hypothesis about the relationship between the natural and vNM measures is natural, this one seems reasonably natural. But it implies that for some increasing and strictly concave function w, $w \circ g$ is the vNM measure. Thus by itself (ii) does not imply that the priority view is inconsistent with Reduction to Prospects, and therefore certainly does not imply the Priority Principle. On the other hand, once we accept (iii) and thus accept the entirety of the standard definition of the priority view, it does follow that the priority view implies the Priority Principle. Thus to that extent, the Priority Principle has been anticipated.[16]

However, two things should be noted. First, on this standard account of the priority view, the Priority Principle is a derived claim. By contrast, I

[16] McKerlie (2001) is sometimes cited as the first person to have argued for what the Priority Principle implies about one-person cases. But this interpretation seems mistaken. It is based on a brief remark on p. 283 of his article, but that is the only place where he mentions risk, so it is at least consistent, and I am inclined to think more natural, to read him as implicitly assuming the natural measure.

am claiming that the Priority Principle should be seen as the fundamental claim of the priority view. Second, the standard account of the priority view, and therefore the derivation of the Priority Principle, presupposes that the individual preorder satisfies the expected utility axioms, and involves quantitative talk about welfare. By contrast, my account of the priority view, and motivation for the Priority Principle as the key component, avoids both those things.

If (PD) did express a natural ethical idea, when understood with respect to the vNM measure, then the axiomatization of the priority view which leads to the standard definition would, in my view, provide a reasonably compelling defence of the priority principle. But I have argued that (PD) fails in that respect, and that is one of my main reasons for rejecting the standard definition of the priority view. Hence the onus is now upon me to try to show that the Priority Principle does express a natural ethical idea.

8.3. Defence

Assuming that the antecedent of (ii) of the Priority Principle is sometimes met, what is so puzzling about the Priority Principle can be crystallized by noting that it implies the following.

> (D) In one-person cases, the individual preorder and the social preorder are not identical. In particular, the social preorder is more risk averse than the individual preorder.

I should say that in the most general characterization of prioritarianism, it is probably not accurate to say that according to prioritarians, the social preorder is more risk averse than the individual preorder. But this is because of a strange possibility for the individual preorder, one which in practice all prioritarians will wish to reject. This will be discussed more at the end of section 9. But the slight inaccuracy in (D) is fairly harmless, and makes it easier to illustrate other ideas.

Call the betterness relation in one-person cases the 'one-person betterness relation' and recall that the individual preorder encodes interpersonal and intrapersonal comparisons.[17] It will often be more vivid to use this terminology, and I will freely switch. For example, (D) becomes the claim that the one-person betterness relation and the sole person's individual betterness relation come apart, with the former being more risk-averse.

[17] Euthyphro-style questions arise: should we see the individual preorder as determined by interpersonal and intrapersonal comparisons, or alternatively, as determining. But we don't need to settle this here.

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

Many writers have accepted that (D), or something similar, is a consequence of the priority view.[18] If (D) can be defended, so can the Priority Principle. But (D) is puzzling to say the least. There have been indirect attempts to defend (D) by first defending (PD) and using that to argue for (D), most notably by Rabinowicz (2001, 2002). But I am aware of almost no direct attempts to defend it, though two ideas are at least hinted at by Parfit (2012).

First, Parfit in effect associates the idea that in one-person cases, the individual and social preorders coincide with a Pareto principle. He then associates the Pareto principle with utilitarianism, and points out that there are alternatives to utilitarianism, contractualism in particular. But it is hard to see how contractualism is of any help. Contractualists stress the importance of unanimity: principles must be acceptable from everyone's point of view (Nagel, 1970). But rather than opposing that contractualist idea, Reduction to Prospects seems to be a direct expression of it. In particular, it is hard to see what principle could be more acceptable to the sole person in one-person cases than the idea that what's better and what's better for her coincide!

Second, for the remainder of this section, assume that a, b and c are distinct welfare levels, ranked from highest to lowest. Suppose that $[\frac{1}{2}, a; \frac{1}{2}, c]$ is a better prospect than b. But there is a one in two chance that b would actually be better than the prospect. In the case Parfit considers, c is a poor welfare level, involving severe harm, and with the qualification made in the footnote, Parfit concludes that the history $[b]$ is better than the lottery $[\frac{1}{2}, a; \frac{1}{2}, c]$.[19] But the assumption that c is a poor welfare level does not help the general defence of (D), where the difference in risk-aversion applies at all welfare levels. So when this assumption is stripped away, Parfit's claim becomes: because c is worse than b and has a one in two chance of resulting, the history $[b]$ is better than the lottery $[\frac{1}{2}, [a]; \frac{1}{2}, [c]]$. But as an argument for (D), this is doubly unconvincing.

The first reason is that it simply does nothing to address, let alone defuse, the tension between the claim that the lottery is better for the sole person than the certain outcome, while the outcome is better than the lottery. Surely something must be said to explain why betterness and betterness for the sole person point in opposite directions. To try to defuse the tension, Parfit does makes a general appeal to a pluralistic view

[18] The first to argue for this was Rabinowicz (2001, 2002), while McCarthy (2008) provided a formal derivation along similar lines. See also Otsuka and Voorhoeve (2009).

[19] See Parfit (2012, sec. V). What Parfit actually says is that the history could be rationally chosen, but the claim of the text is what is needed to defend (D) or the Priority Principle. However, I find Parfit's discussion difficult to follow, and I may be misrepresenting him. It is possible that in the relevant passage, he is has already accepted (D) on the familiar grounds of first accepting (PD), understood with respect to the vNM measure, and is simply discussing or endorsing its implications.

about ethics he attributes to Sidgwick. If pluralism is true, different ethical factors sometimes pull in different directions. But just because we grant the fairly banal claim that some ethical factors pull in different directions in some contexts (what is better for Ann may be worse for Bob), it does not follow that all ethical factors pull in different directions in all contexts. In particular, accepting a modest version of pluralism hardly commits us to the claim that betterness and betterness for Ann pull in different directions when she is the only person around. Pluralism, by itself, provides no defence of (D).

The second reason is that there is no explanation of what is wrong with reversing the argument. Suppose instead that b is a better prospect than $[\frac{1}{2}, a; \frac{1}{2}, c]$. Why is the reverse of Parfit's claim not just as convincing: because a is better than b and has a one in two chance of resulting, the lottery $[\frac{1}{2}, [a]; \frac{1}{2}, [c]]$ is better than the history $[b]$? Thus the difficulty with Parfit's example is that it gives us no reason to break symmetry between individual and one-person betterness in favour of the latter being more risk-averse, which is what we seek, rather than less risk-averse.

Recall that Rabinowicz (2001, 2002) suggested that the priority view is motivated by Rawls's separateness of persons objection to the impartial spectator argument. Earlier I claimed that this motivation was unsuccessful. Nevertheless, I am now going to suggest that paying attention to Rawls's objection does lead to the kind of symmetry-breaker we need. I will later raise some doubts about this argument, but it is the best I can think of. But first we need to review Rawls's objection.

8.4. The impartial spectator argument

According to this argument, two histories can be compared as follows (Nagel 1970; Rawls 1971). The lives of each member of the population in the first history are concatenated into a single life via something like the serial-life construction of Lewis (1946), with amnesia in between each component life; likewise for the second history. The impartial spectator imagines leading the lives in full detail, one after the other, then ranks the histories by ranking their associated serial-lives.

Rawls (1971) famously objected that this argument ignores the separateness of persons. Talk of compensation is appropriate within a single life because the person who is leading it can be compensated for suffering a burden at one time by receiving some benefit at another time. But talk of compensation is not appropriate when distinct persons are involved, since one individual is not compensated for suffering some burden by some other individual gaining some benefit. Therefore, Rawls concluded, there are more restrictions on transfers of benefits when talk of compensation is not appropriate than when it is.

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

In some ways, Rawls overstates his objection: it is directed at the impartial spectator argument, but he immediately concludes that it applies to all forms of utilitarianism. But without going into details, the impartial spectator argument presupposes a welfare measure which is not obviously the vNM measure. So it is unclear what relevance Rawls's objection has to an argument for utilitarianism understood in terms of the vNM measure which appeals to a variant of Harsanyi's aggregation theorem. Nevertheless, many writers have found Rawls's claims about the significance of compensation appealing.

8.5. The individual preorder

A huge body of literature is devoted to searching for reasonable constraints which will determine the social preorder given the individual preorder. When philosophers think of 'the problem of aggregation', this is typical of the kind of problem they have in mind. The individual preorder is thought of as given for free, and the difficulty is then to determine the social preorder. Nevertheless, determining the individual preorder itself involves addressing significant problems of aggregation.

To illustrate, suppose we have to make some sort of evaluative judgement (comparative or absolute) about the history $[a, c]$. To arrive at such a judgement, we have to take into account the welfare levels of the two different people. But now suppose we wish to make some sort of evaluative judgement (comparative or absolute) about the prospect $[\frac{1}{2}, a; \frac{1}{2}, c]$. To do this we have to take into account the welfare levels in the two different states of nature. Thus the second problem involves no less of a problem of aggregation than the first: each requires us to take into account two locations of welfare to arrive at an overall judgement. This gives rise to what I believe is a much neglected problem: what determines the content of the individual preorder?

It's worth noting how two lines of thought do not provide much of an answer. First, philosophers have had a lot to say about welfare; thus there are several competing accounts of what it amounts to. But learning, for example, that welfare is happiness tells us very little about how to order prospects over levels of welfare. For example, if a corresponds to 'very happy', b to 'mildly unhappy' and c to 'very unhappy', we are still in the dark about how to compare $[\frac{1}{2}, a; \frac{1}{2}, c]$ with b. Second, I suspect that welfare economists have downplayed the challenge of determining the individual preorder by assuming that individual preferences over prospects, most likely those formed in ideal circumstances, will somehow settle the matter. But even if we grant the far-fetched assumption that there is some reasonably determinate fact of the matter about what people would prefer in some non-question begging characterization of ideal circumstances, many philosophers believe that people can easily prefer

what is worse for themselves no matter what kinds of cognitive filters are imposed on their preferences. But then there is no obvious reason to assume that people's preferences are automatically right about risk. So what instead could help determine the content of the individual preorder? It is a hugely important problem.

However, notice that to defend (D), we do not need to fully solve the problem. We merely need to find some factor which is, say, relevant to determining the individual preorder but not the one-person social preorder, and we want this symmetry-breaker to justify the one-person social preorder being more risk averse than the individual preorder.

8.6. Breaking symmetry

Suppose an individual A has to choose between welfare level b for certain versus the prospect $[\frac{1}{2}, a; \frac{1}{2}, c]$. In many cases, it will be relatively uncontroversial to say that $[\frac{1}{2}, a; \frac{1}{2}, c]$ is better for her than b. But why is this, given that under the prospect, she would face a one in two chance of ending up with c, which is worse for her than b? The obvious response is that under the prospect, she would also face a one in two chance of ending up with a, and in this case the risk of gain, relative to the baseline of b, more than compensates her for the risk of loss. Thus in thinking about the content of her individual betterness relation, talk of compensation is entirely appropriate, even when the benefits and burdens are in different states of nature. It is the fact that in prospect the individual herself stands both to suffer the harm and enjoy the benefit which makes talk of compensation appropriate. Unlike the context Rawls was discussing, only one person is involved.

However, from another perspective, talk of compensation is inappropriate. Suppose A faces the lottery $[\frac{1}{2}, [a]; \frac{1}{2}, [c]]$, and that this lottery is operationalized by the toss of a fair coin, with heads corresponding to $[a]$ and tails $[c]$. Tails comes up and A ends up with the lowest level of welfare, c. But it seems plausible that the truth of the counterfactual, 'Had heads come up, A would have had the highest level of welfare a', makes no difference to what her actual life is like. It is not like the case where she transfers a benefit from one part of her life to another so that her actual life is on balance just as good.

The idea that what A's life would have been like had heads come up has no evaluative bearing on what it is like given that tails comes up is related to the various independence axioms of expected utility theory. The axiom Strong Independence defined earlier is one such axiom. The motivation for Strong Independence was expressed by Samuelson (1952): 'Either heads or tails must come up: if one comes up, the other cannot.' Or as Broome (1991) puts it, there are no evaluative interaction effects between states of nature.

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

To make this more concrete, compare the following lotteries involving a fair coin and three different people, each of which will result in a single person coming to exist: L = [heads, [A has a]; tails, [A has c]], and M = [heads, [B has a]; tails, [C has c]]. Extending the idea of anonymity to cover cases involving different people in the obvious way, the heads outcomes are clearly equally good, and the tails outcomes are also equally good. Given that there are no evaluative interaction effects between states of nature in such a case, the fact that L involves the same individual in both outcomes has no relevance, and we should conclude that L and M are equally good.[20]

Thus the following argument is now available to us. (i) The individual preorder encodes, among other things, individual betterness relations. (ii) In determining the content of individual betterness relations, talk of compensation is appropriate, and plays an explanatory or justificatory role. (iii) But on a natural way of determining one-person betterness relations, talk of compensation is irrelevant. (iv) But according to Rawls and a view many philosophers find appealing, there are more restrictions on transfers of benefits when talk of compensation is not appropriate than when it is. (v) Therefore, individual betterness relations are less averse to transfers across states of nature than one-person betterness relations. Example: if b and $[\frac{1}{2}, a; \frac{1}{2}, c]$ are equally good prospects, the history $[b]$ is better than the lottery $[\frac{1}{2}, [a]; \frac{1}{2}, [c]]$. Thus in such cases, the social preorder is more risk averse than the individual preorder, as required.

I have my doubts about how good a rationale this is for (D), or more generally, the Priority Principle. But it is the best defence I can think of. Section 11 will examine some criticisms, but I end the defence with an autobiographical remark. I find that when I think prospectively about financial investments, I focus more on expected returns and less about risk: the expected returns are mine, and this compensates me for the risk. But when I think instead about my various possible future selves in different states of nature as people I am responsible for, and ask which investment is best, I focus less on expected returns, and more on reducing risk. This is merely a psychological observation, but if it is widely shared, it suggests that the asymmetry between individual betterness relations and one-person betterness relations built into the Priority Principle is fairly natural.

[20] This argument requires a much weaker independence idea than the one built into Strong Independence, so we can accept this argument without being committed to the claim that the social preorder satisfies Strong Independence. More precisely, treat non-existence as a welfare level and extend Two-Stage Anonymity in the natural way to allow the population to vary, as in McCarthy *et al.* (2016). This then implies that L is exactly as good as M.

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

9. THE CORE OF THE PRIORITY VIEW

McCarthy *et al.* (2016) claim that the core of utilitarianism is what we call[21]

> **Generalized utilitarianism** Anteriority, Two-Stage Anonymity and Reduction to Prospects.

In this article I am proposing that the core of the priority view is what I will call

> **Generalized prioritarianism** Anteriority, Two-Stage Anonymity and the Priority Principle.

Generalized utilitarianism is a substantial generalization of Harsanyi's utilitarianism. Many of the advantages of this generalization discussed in McCarthy *et al.* (2016) apply to generalized prioritarianism. I will not try to rehearse them all, but two things are worth noting: (i) generalized prioritarianism assumes none of the expected utility axioms, either for the individual or social preorder; and (ii) the motivation for and characterization of generalized prioritarianism makes no use of any quantitative evaluative measures, welfare measures in particular.

This section provides a representation theorem for generalized prioritarianism, and then relates it to the way prioritarianism has traditionally been defined. For those who wish to skip technicalities, the upshot is that generalized prioritarianism turns out to admit a simple characterization, and when the assumptions underlying the standard definition of prioritarianism are added, we end up recovering that definition.

To characterize generalized prioritarianism more precisely, we need some terminology. Let X be an arbitrary set (thought of as a set of possible outcomes) and $\mathcal{P}(X)$ be the set of probability functions on it. Let $\succsim_1$ be a preorder on $\mathcal{P}(X)$. Say that a preorder $\succsim_2$ on $\mathcal{P}(X)$ is *outcome-equivalent* to $\succsim_1$ if (identifying each $x \in X$ with the probability function which gives x a probability of one): for all $x, y \in X$, $x \succsim_1 y \iff x \succsim_2 y$. To illustrate, in the case where X is the set of welfare levels and $\mathcal{P}(X)$ are prospects, this just means that $\succsim_1$ and $\succsim_2$ agree on the way they rank welfare levels. Say that $\succsim_2$ is a *strictly concave transform* of $\succsim_1$ if for all x, y, and $z \in X$ with $x \succ_1 y \succ_1 z$: $y \sim_1 \frac{1}{2}x + \frac{1}{2}z \Rightarrow y \succ_2 \frac{1}{2}x + \frac{1}{2}z$.

For any $p \in \mathcal{P}(X)$ define the expectation operator E_p as follows. For any function $f : X \to \mathbb{R}$, $E_p(f) := p_1 f(x_1) + \ldots + p_m f(x_m)$ where $p = [p_1, x_1; \ldots; p_m, x_m]$. Thus E_p just computes the expected value of f relative to the probability function p.

Recall from section 2 that for any lottery L, $\mathcal{P}_i(L)$ is the prospect individual i faces under L, and that we are assuming that the population

[21] This is for the constant population case; we characterize the more general variable population case along very similar lines.

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

is $\{1, \ldots, n\}$. Let p_L be the prospect defined by

$$p_L := \tfrac{1}{n}\mathcal{P}_1(L) + \ldots + \tfrac{1}{n}\mathcal{P}_n(L).$$

In other words, p_L is the prospect one would face if one were behind a veil of ignorance with an equal chance of being any member of the population facing L. Let $\mathbb{W}$ be the set of welfare levels, $\mathbb{P}$ the set of prospects, $\mathbb{H}$ the set of histories and $\mathbb{L}$ the set of lotteries. Then with the domain assumptions given in section 2, the following is a special case of McCarthy *et al.* (2016, Theorem 1).

Theorem 1 *Given a preorder $\succsim_{\mathbb{P}}$ on $\mathbb{P}$, the following two claims are equivalent.*
(a) Generalized utilitarianism.
(b) For all L, L' *in* $\mathbb{L}$,

$$L \succsim L' \iff p_L \succsim_{\mathbb{P}} p_{L'}.$$

The Appendix shows that this straightforwardly entails the following.

Theorem 2 *Given a preorder $\succsim_{\mathbb{P}}$ on $\mathbb{P}$, the following two claims are equivalent.*
(a) Generalized prioritarianism.
(b) There exists a preorder $\succsim_{PV}$ on $\mathbb{P}$ which is an outcome-equivalent strictly concave transform of $\succsim_{\mathbb{P}}$ such that for all L, L' *in* $\mathbb{L}$,

$$L \succsim L' \iff p_L \succsim_{PV} p_{L'}.$$

Specializing to a more familiar environment, the Appendix proves the following.

Corollary 1 *Suppose the individual and social preorders satisfy the expected utility axioms. Let $u : \mathbb{W} \to \mathbb{R}$ be such that for any prospects p and p', $p \succsim_{\mathbb{P}} p' \iff E_p(u) \geq E_{p'}(u)$. Suppose $I := u(\mathbb{W})$ is an interval of real numbers of positive length. Then the following two claims are equivalent.*
(a) Generalized prioritarianism.
(b) There exists an increasing and strictly concave function $w : I \to \mathbb{R}$ such that for all L *and* L' *in* $\mathbb{L}$,

$$L \succsim L' \iff E_L(W) \geq E_{L'}(W)$$

where $W : \mathbb{H} \to \mathbb{R}$ is defined by $W(h) := w(u(x_1)) + \ldots + w(u(x_n))$ for any history $h = [x_1, \ldots, x_n]$.

Thus if we assume (i) the expected utility axioms for the individual and social preorder; (ii) the domain is well-behaved, and (iii) the vNM welfare measure, generalized prioritarianism yields the standard definition of prioritarianism given in section 8.2.

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

This is not the place to explore a technical development of the theory,[22] and I will merely make three brief remarks.

First, we have been working with a fixed population. But nothing in our official characterization of generalized prioritarianism implies that the social preorders relative to one population are effectively the same as the social preorders relative to another population. For example, if $\succsim_2$ is the social preorder for some population of two people and $\succsim_3$ is the social preorder for some population of three people, and if a, b and c are welfare levels, it does not follow that $[b,b] \succsim_2 [\frac{1}{2},[a,a];\frac{1}{2},[c,c]] \iff [b,b,b] \succsim_3 [\frac{1}{2},[a,a,a];\frac{1}{2},[c,c,c]]$. But this implication is obviously desirable for prioritarians, and it holds if we accept a natural generalization. To state it, observe that if some lottery L involving some fixed population guarantees equality, p_L is the prospect faced by each member of the population under L.

> **Uniformity** Let $\mathbb{I}_1$ and $\mathbb{I}_2$ be any fixed populations with respective social preorders $\succsim_1$ and $\succsim_2$. Let L and L' be any lotteries involving $\mathbb{I}_1$, and M and M' be any lotteries involving $\mathbb{I}_2$, such that (i) L, L', M and M' all guarantee equality of outcome, and (ii) the prospect members of $\mathbb{I}_1$ face under L is identical to the prospect members of $\mathbb{I}_2$ face under L', and the prospect members of $\mathbb{I}_1$ face under M is identical to the prospect members of $\mathbb{I}_2$ face under M'. Then $L \succsim_1 L' \iff M \succsim_2 M'$.

If we add this axiom to generalized prioritarianism, when it yields the standard definition of prioritarianism in the way just noted, the result is that social preorders for different populations are all governed by the same strictly concave function w.

Second, generalized prioritarianism is a very weak theory, and is compatible with many different accounts of both the structure and content of the relation $\succsim_{PV}$ in Theorem 2. Any sane prioritarian will want to further constrain this relation, directly or indirectly, but it is an advantage of the approach that it can be constrained in different ways. For example, some may wish to impose all of the expected utility axioms, others may not. In addition, for applications, one may wish to ensure that the individual preorder is rich enough for the antecedent of the Priority Principle to hold sufficiently often.

Third, there is a case for stating the Priority Principle more generally as follows: for any welfare level x and non-degenerate prospect p such that x is as good a prospect as p, with x worse than at least some of the welfare levels which could result from p, the history $[x,\ldots,x]$ is better than the

[22] Many results about prioritarianism can be obtained by mimicking the results in McCarthy *et al.* (2016) about utilitarianism. These include results for general probability measures, various styles of vector valued expected utility-style representations, nonexpected utility representations, and so on.

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

lottery L which guarantees equality and in which every individual faces p. But without further constraints on the individual preorder and its relation to the social preorder, I have been reluctant to extend the Priority Principle beyond the simplest possible case.

As already noted, it is tempting to say that the Priority Principle expresses a notion of social risk aversion. Now there is a strange case which suggests that the Priority Principle is not a complete expression of the basic prioritarian idea, and which also shows that it is a mistake to say that the basic prioritarian idea expresses a notion of social risk aversion.[23] But this kind of case can be ruled out by assuming that the individual preorder satisfies a condition known as stochastic dominance, which is a very weak and uncontroversial condition. With that assumption in place, it does seem natural to see the Priority Principle as expressing a notion of social risk aversion. There is a vast body of literature connecting risk aversion with inequality aversion, so this may yield a profitable way of drawing conceptual connections between prioritarianism and egalitarianism, and perhaps generalizing or specializing the Priority Principle. But this topic will have to wait for another occasion.

10. TAXONOMY

McCarthy *et al.* (2016) effectively identify utilitarianism with generalized utilitarianism plus the assumption that the social preorder satisfies Strong Independence.[24] I will likewise identify prioritarianism with generalized prioritarianism plus the same assumption. For familiarity, I will now just talk in terms of utilitarianism and prioritarianism, but the sequel applies just as much to their generalized versions.

On the taxonomy used here, there are three fundamental principles, Anteriority, Two-Stage Anonymity and Reduction to Prospects, and consequently three basic choices. The first is whether to accept or reject Anteriority. The classic example of a theory which rejects Anteriority but accepts the other two is *ex post* egalitarianism, though there are many others. The second is whether to accept or reject Two-Stage Anonymity. The classic example of a theory which rejects Two-Stage Anonymity but accepts the other two is *ex ante* egalitarianism. The third is whether to accept or reject Reduction to Prospects. I have been arguing that

[23] In more detail, suppose $y \succ_{\mathbb{P}} x$, $y \succ_{\mathbb{P}} z$ and $y \sim_{\mathbb{P}} [\frac{1}{2}, x; \frac{1}{2}, z]$. It seems natural to say that giving more weight to the worse off than is given by Reduction to Prospects (and hence generalized utilitarianism) should imply that $[\frac{1}{2}, [x]; \frac{1}{2}, [z]] \succ [y]$. Thus if anything, prioritarianism should be seen as risk-seeking in this case. Of course, this case is very strange, and it is the kind of thing which is ruled out by requiring the individual preorder to satisfy stochastic dominance.

[24] More precisely: we define utilitarianism as generalized utilitarianism plus Strong Independence for the individual preorder, but that is equivalent to the claim in the text.

prioritarianism rejects Reduction to Prospects, but accepts the other two. I will call theories which, like prioritarianism, reject Reduction to Prospects, *non-reductive*, and theories which accept it *reductive*.

Since there are three independent binary choices, there are eight possible types of theories, which can be illustrated by hybrids of the theories just mentioned. For example, one can hybridize *ex post* and *ex ante* egalitarianism to get a theory which rejects Anteriority and Two-Stage Anonymity. Likewise, one can hybridize *ex post* egalitarianism and prioritarianism to get a theory which rejects Anteriority and Reduction to Prospects. I do not have the space to go into any detail about how to do this, but the basic trick to formulate the latter hybrid is to modify the *ex post* egalitarian axioms by replacing reference to the individual preorder with reference to an outcome-equivalent strictly concave transform of the individual preorder.

I believe McCarthy (2006, 2008) has the dubious distinction of introducing the terms '*ex ante* prioritarianism' and '*ex post* prioritarianism.' This terminology seems to have become established, but I now believe that it is a serious mistake. Prioritarianism is not a theory which has two different forms, '*ex ante*' and '*ex post*'. There is just a single theory, prioritarianism, which accepts Anteriority and Two-Stage Anonymity. What I called *ex ante* prioritarianism did not deserve the name, and was simply a special case of *ex ante* egalitarianism. Since there is no theory which deserves to be called *ex ante* prioritarianism,[25] it is pointless to call prioritarianism '*ex post* prioritarianism'.

Parfit (1991) gives the impression that (ex post) egalitarianism and prioritarianism are in many ways very similar, and reach the same conclusions about a wide range of cases. Utilitarianism is more or less dismissed without discussion, so the real issue becomes the infamous 'equality or priority?' question. A large body of literature has followed suit. However, while it is true that in risk-free cases, *ex post* egalitarianism, *ex ante* egalitarianism and prioritarianism can be made to look very similar (McCarthy, 2015), this resemblance is quite superficial. On the view I have been developing, their disagreements are essentially about risk, so these disagreements do not show up in risk-free cases.

In particular, these three theories depart from utilitarianism in quite different ways. To illustrate, let us compare prioritarianism with *ex post* egalitarianism, and thus take Two-Stage Anonymity for granted. Prioritarianism accepts Anteriority but rejects Reduction to Prospects. Egalitarianism accepts Reduction to Prospects but rejects Anteriority. But accepting Anteriority surely does not provide an ounce of support for rejecting Reduction to Prospects. Likewise, accepting Reduction to

[25] Of course, there is a theory which hybridizes prioritarianism and *ex ante* egalitarianism, but it is highly misleading to call it '*ex ante* prioritarianism'.

Prospects provides no support for rejecting Anteriority. It is therefore strange to say the least that so much literature assumes that reasonable credence is close to one on 'equality or priority'.

11. THE TRUTH IN THE PRIORITY VIEW

I now consider two types of objection to the priority view, and end with a suggestion concerning what exactly is true about the priority view.

11.1. Ethical objections to the priority view

Recall that the levelling-down objection against egalitarianism discussed in section 6.3 objected that given a continuity assumption,[26] (*ex post*) egalitarians are committed to the existence of cases where it is on-balance better to make everyone strictly worse off in order to reduce inequality. But given that continuity assumption, prioritarians are committed to the existence of cases where it is on-balance better to make everyone strictly worse off in order to give priority to the worse off. Thus both egalitarianism and prioritarianism are committed to the following.

> (W) It is sometimes on-balance better to make everyone face worse prospects to achieve some social goal.

Although Parfit's original levelling-down objection is connected with (W), it is not the original levelling-down objection itself which shows that egalitarianism is objectionable in my view, but rather the commitment to (W). Thus although it is doubtless wrong to say that prioritarianism is subject to a levelling-down objection, it is committed to (W). Thus egalitarianism and prioritarianism seem roughly on a par in having an appearance of prima facie implausibility. In particular, it is surprising that some writers have taken a commitment to (W) to be a decisive objection to prioritarianism, and have immediately concluded that egalitarianism is a safe harbour.

Of course, the prima facie implausibility of one or both of egalitarianism and prioritarianism could be nevertheless be overturned by a suitable defence. In one respect egalitarianism has the advantage. Myerson's example in section 6.1 encapsulates what egalitarianism is all about in a way which makes its appeal manifest, but there is no such parallel example for prioritarianism. However, once one notes that (W) follows in the wake of egalitarianism, there is a serious question about what the rationale for egalitarianism is, and in my view, at least, there is not a successful one (McCarthy, 2015). Since we have at least some sort

[26] I will take this assumption for granted in what follows. This makes little real difference, but helps to make the discussion more vivid.

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

of defence of prioritarianism, it is possible that prioritarianism comes out ahead. But that defence was rather complicated, and deserves scrutiny.

Let a, b and c be welfare levels ranked from highest to lowest, and suppose facing $[\frac{1}{2}, a; \frac{1}{2}, c]$ is exactly as good for A as having b for certain. It is banal to claim that in facing the prospect the chance of gain, relative to the baseline of b, exactly compensates A for the risk of loss. Thus the crucial step (ii) of the defence of the priority view is surely correct to say that talk of compensation is appropriate in such cases. But it goes on to infer that because it is appropriate, it is somehow playing a role in determining the content of A's individual betterness relation. If that inference is dropped, the defence of the Priority Principle in section 8.6 does not work.

But the inference appears to be based on a sleight of hand. It is natural to use talk of compensation to report the content of A's individual betterness relation. Thus for any welfare levels x, y and z ranked from highest to lowest, in comparison with y for certain, A is compensated for the risk of loss in facing $[\frac{1}{2}, x; \frac{1}{2}, z]$ just in case facing $[\frac{1}{2}, x; \frac{1}{2}, z]$ is exactly as good for her as having y for certain. But this does not imply that anything to do with compensation has any role in determining the content of A's individual betterness relation. Reporting does not imply determining. Friends of the priority view need an argument to close this gap, and I am afraid I cannot see one.

I mentioned a personal psychological observation in the final paragraph of section 8.6. If that observation applies to others, there would seem to be tacit support for the Priority Principle. But in the absence of a way of rescuing the argument just mentioned, it seems much more natural to regard that observation as merely reporting a framing effect.

11.2. Conceptual objections to the priority view

It is tempting to think that the plausibility of the priority view depends on whether the Priority Principle is an ethically plausible idea. But there is a different worry about the priority view.[27]

Prioritarians, and also those who argue against the priority view on substantive ethical grounds, accept

> (X) It is a substantive ethical question what the relation is between the individual preorder and the one-person social preorder.

I will argue instead that we ought to accept

[27] The idea that the priority view is in some sense a conceptual illusion has been defended in Broome (1991), McCarthy (2008), and Greaves (2015), and is implied by Hammond (1991). As far as I am aware, no one who holds the position this view is directed against has genuinely responded, with the exception of Adler (2011).

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

(Y) It is a conceptual truth that the individual preorder and the one-person social preorder coincide.

An immediate objection to (Y) is that it cannot be correct because it obviously makes sense to ask what the relation is between the individual preorder and one-person social preorder. But in response, first, I am not claiming that 'the individual preorder and one-person social preorder coincide' is true as a matter of meaning; and second, I do not say that (Y) is true; rather, I am going to argue that we ought to accept it.

To draw a grandiose analogy, much mathematical progress is made by the acceptance and revision of definitions. For example, whoever first proposed the now standard definition of the real numbers was obviously not making a claim about the meaning of 'the real numbers', and was not saying that it is a conceptual truth that the real numbers are such and such (rather than a derived truth from some other definition). What that person was doing was in effect suggesting to the mathematical community that his or her definition ought to be accepted on something like the grounds that it best reflected or promoted evolving mathematical practice. Although incomparably less significant, I am likewise proposing that we should adopt (Y).

I will shortly argue that both (X) and (Y) have some degree of at least tacit support among moral philosophers, so we are already dealing with a tacitly contested topic. To see which of (X) and (Y) we ought to side with, we need to ask what theoretical life is like should we adopt either one of them.

The contest between (X) and (Y) is quite general: (X) implies that it is a substantive ethical question whether a variety of non-reductive theories are true; and (Y) implies that non-reductive theories are false on conceptual grounds. But I do not have the space to develop the framework needed to discuss non-reductive and reductive theories in general, and will focus the debate around prioritarianism and utilitarianism, and begin with five preliminary points.

(i) *No privileged access.* For both utilitarians and prioritarians, determining the content of the individual preorder is a major theoretical problem involving serious questions about aggregation, as already noted in section 8.5. Thus it cannot be said in favour of (X) that we have some kind of privileged epistemic access to the individual preorder but not to the one-person social preorder.

(ii) *The individual preorder is an idle wheel.* For prioritarians, the individual preorder is curiously epiphenomenal. Neither the structure nor the content of the individual preorder imposes much constraint on the social preorder. For example, just because the individual preorder satisfies Strong Independence, it does not follow from generalized prioritarianism

that the one-person social preorder does as well. Likewise, if one thinks of prioritarianism as giving extra weight to lower levels of welfare, the individual preorder provides very little constraint on that weighting.

(iii) *Gratuitous complexity.* Theorems 1 and 2 imply that every social preorder which is prioritarian relative to one account of the individual preorder is utilitarian according to another account of the individual preorder. Thus in one sense, the debate between utilitarians and prioritarians is empty: any social preorder which can be defended by a prioritarian can also be defended by a utilitarian. Thus admitting prioritarianism into our taxonomy of distributive theories does not open up some genuine new possibility for the social preorder which was not already there. However, a prioritarian interpretation of any particular social preorder is always more complicated than the utilitarian interpretation. The utilitarian interpretation connects individual and social preorders in a simple way, whereas the prioritarian connection is more complicated.

(iv) *More gratuitous complexity.* Suppose utilitarians and prioritarians alike start by asking themselves what the one-person social preorder is. Once they have figured that out, they're done as far as determining the content of the social preorder,[28] and for utilitarians, the individual preorder is determined as well. But for prioritarians there is one more task, for it is still an open question for them what the individual preorder is. This further illustrates the gratuitous complexity of prioritarianism. Of course, prioritarians might have started by asking what the individual preorder is. But given how disconnected this is from the social preorder, as noted in (ii) above, it is far from clear what the point of that exercise is.

(v) *A puzzle about meaning.* There is a real puzzle about what prioritarians mean by 'better for a person than'. I am of course not about to offer an analysis of the meaning of such terms as 'better than' and 'better for a person than'. But there is a serious difficulty understanding prioritarians. For example, when Parfit (2012, 404) gave his definition of which way of acting is best according to prioritarianism, he immediately tacks onto the end of the definition: 'If other things are equal, we ought to act in this way.' Parfit clearly thinks, as it is very natural to think, that there is a deep conceptual connection between what is best and what we ought to do. The 'other things being equal' clause is there to allow that the connection might not be as act consequentialists think it is. For example, we could be rule consequentialists, or we could deny that someone ought to do what's best when the cost to herself is excessive. But in a like manner,

[28] At least given Uniformity, for prioritarians, which I am taking for granted. For utilitarians, Uniformity is already built into Reduction to Prospects.

it seems very natural to say that in one-person cases, what is best for the sole person is, other things being equal, what she ought to do. If not, it is very hard to know what role 'best for the person' is playing in our network of evaluative and normative terms. I believe this explains why the most common initial reaction I have met to the prioritarian treatment of one-person cases is the famous incredulous stare: it is just hard to understand what someone who says that such and such is best for the sole person, but is nevertheless worst, is really saying.[29] The prevalence of this reaction is why I believe that (Y) has considerable tacit support, as well as some amount of explicit support.

In summary, the dispute between (X) and (Y) is tacitly contested. This precludes the strategy of winning by stipulating. Thus Adler (2011, 343–4) claims to have answered Broome (1991)'s challenge to prioritarians to draw the principled distinction they need by asserting, in effect, that the individual preorder is to be identified with, or at least constructed from, the preferences of ideally rational individuals. He then asserts (PD) on substantive ethical grounds, leading to a clash between the individual preorder and the one-person social preorder. But this is to beg the question against (Y), not argue against it. What's more, there is a serious question about the coherence of such positions.

Ideal preference accounts claim that in suitable circumstances, if an ideally rational agent weakly prefers a situation s_1 to a situation s_2, then s_1 is at least as good for her as s_2. The 'suitable circumstances' clause is there because an ideally rational agent could easily prefer one situation to another despite the fact that it is worse for herself, simply because it is better for others. To avoid circularity, there should be some kind of independent specification of the suitable circumstances, and it seems very natural to identify them with circumstances in which no other agents are around to care about. But if our ideally rational agent is also a prioritarian, in these one-person worlds, she can still end up preferring one situation to another despite the fact that it is worse for herself. This suggests that the workability of ideal preference accounts may well depend upon accepting (Y).

I am not going to develop this argument in any detail as many people reject ideal preference accounts. But it does suggest that the dispute between (X) and (Y) is surprisingly deep, and resistant to easy resolution. Thus we are back to having to assess the theoretical costs and benefits of each.

I cannot see any way in which theoretical life is easier if (X) is true. For example, (X) yields no accounts of the social preorder which cannot also be advocated given (Y). In addition, theoretical life is harder in several ways if (X) is true: it is more complicated; the content and structure of the

[29] See section 11.2.1 below for some minor qualifications.

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

individual preorder is to a large degree irrelevant to the social preorder; and it involves a strange account of the meaning of 'better for a person than'. On some views, this is enough to make (Y) true. But it is less controversial to say that that (X) and (Y) are two competing conventions in the philosophical community, and that we ought to accept (Y).

An analogy may help. According to Lewis (1973), there are many theories which provide true, though possibly partial, descriptions of the universe. Some of these are simpler than others, and some are stronger than others, in the sense of being more informative. The laws of nature are the theorems of the theory which does best overall in terms of the competing virtues of simplicity and strength. Suppose there is such a theory, and compare it with another theory which does equally well in terms of strength, but worse in terms of simplicity. What makes the latter false as an account of the laws of nature is not getting the facts wrong, but being too complicated. Similarly, what makes prioritarianism false, on the present objection, is not getting the content of the betterness relation wrong, for utilitarianism and prioritarianism can perfectly well agree about that; what makes it false is its gratuitous complexity.

Finally, the theme of simplicity leads to a tu quoque response to the writers who in practice are tacitly committed to (X), namely prioritarians and those who argue against prioritarianism on substantive ethical grounds. As already noted, recent participants in this dispute have adopted the vNM measure, or what Broome (1991) called Bernoulli's hypothesis, largely without comment. Broome's argument for Bernoulli's hypothesis was that prima facie, it is indeterminate how to measure welfare, but Bernoulli's hypothesis provides a simpler and more natural way of measuring welfare than other approaches. Broome (1991) claimed that this was enough to show that Bernoulli's hypothesis was true as a matter of meaning. But that is overly strong, and Broome (2004) retreated to the more modest claim that it would be useful to resolve the indeterminacy about welfare measures, and he recommended Bernoulli's hypothesis as a good resolution, I assume still on the grounds of simplicity and naturalness (McCarthy 2007).

But if that is the rationale for Bernoulli's hypothesis, and as far as I am aware, prioritarians and their substantive opponents have provided no other, the same style of rationale also supports (Y). Thus the position held by recent prioritarians and their substantive opponents is somewhat incoherent. I have been framing this as a debate about utilitarianism versus prioritarianism, but these are really stand-ins for reductive versus non-reductive theories. Thus I am claiming that we ought to reject non-reductive theories on conceptual grounds.[30] Such theories include

[30] Parfit (2012, 403) seems to be discussing something along the lines of the conceptual objection to prioritarianism when he says that it would also apply to egalitarianism. But

hybrids of egalitarianism and prioritarianism, sufficiency versions of prioritarianism which only give greater weight to the worse off below a certain wellbeing level, and so on.

11.2.1. Objections. I end with a potential objection to (Y), based on Broome (2004, 92–94). Broome's target is his earlier argument for Bernoulli's hypothesis in Broome (1991) which he thought showed that the priority view is untenable as a matter of mere meaning (Broome 1991: ch. 10, esp. 220–222). There are many complications in Broome's earlier and later discussions which I am afraid I am going to have to bypass (but see Jensen 1995), and I will just try to give the flavour of his later view.

Broome characterizes the basis of his earlier argument as accepting something along the lines of

(G) Personal goodness is determined by general goodness.

For example, how good a life is for a person is determined by how good it is that she has that life. The basis of his later rejection of (G) is that it makes sense to assert ethical positions according to which (G) is false (Broome 2004: 92). Since (G) is supposed to be a view about meaning, the fact that positions which conflict with it make sense is supposed to show that (G) must be wrong, and that suggests that (Y) is also wrong.

An initial response is that the objection holds the target to too high a standard. (Y) is being proposed as an unobvious conceptual truth, or more weakly and more precisely, as something which we ought to adopt as a conceptual truth. The claim that there are ethical positions which make sense and which clash with (Y) by itself need not trouble the proposal. Making sense is too easy; many positions which clash with Lewis's analysis of laws of nature make sense, in that we find them intelligible, understand their implications, and so forth. But that does not by itself refute Lewis's position, which I take to be a form of conceptual analysis. What would refute it would be a demonstration that its theoretical costs outweigh its benefits. With this in mind, let us look at two specific examples from Broome which may seem to put pressure on (Y).

The first is the view that even in one-person cases, it makes conceptual sense to believe in temporal discounting, and think that it is better that the same life be lived earlier rather than later, even though in either case, the welfare levels will be the same (Broome 2004: 92–3). This puts pressure on (Y) since the mere time at which a life is lived is irrelevant to the individual preorder, which is concerned with what lives are like from the inside, but relevant, on the present view, to the one-person social preorder. Nevertheless, if we want to accommodate this view, we can go

that is a mistake. The conceptual objection is only to non-reductive theories, not reductive theories like egalitarianism.

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

beyond the framework of this article and characterize one-person social preorders as holding between something like temporally indexed welfare levels (McCarthy 2008: 22–23). But then we can just modify the letter of (Y) and preserve the spirit to get something like:

> (Y*) It is a conceptual truth that the individual preorder and the one-person social preorder coincide, when the latter is restricted to comparisons between welfare levels with identical temporal indices.

This is obviously of no help to the defender of prioritarianism, and it seems plausible than this kind of trick could handle other potential counterexamples in this vein. At least, the onus is now on the opponent of (Y) to show otherwise.

The second example is based on a distinction often drawn in the literature on population ethics between lives worth living and lives worth creating. For example, part of a common strategy to avoid the so-called repugnant conclusion of Parfit (1984) is to claim that there are lives worth living (often described as lives 'barely worth living') which are nevertheless not worth creating. This seems to be a paradigmatic instance of a distinction between personal and general goodness, and may thus seem to challenge (G). But this is no challenge to (Y), for the two relations it is concerned with are silent on comparisons involving nonexistence, and I am not claiming (or denying) that (Y) remains true when the two relations are extended to such comparisons.

In summary, (Y) is quite a specific claim. It can avoid two natural objections to the more general (G) either by a principled refinement, or by pointing to its limited scope. What prioritarians and their substantive opponents need is to argue that the balance of theoretical benefits over costs in the contest between (X) and (Y) comes out in favour of (X). But as far as I can see, all the advantages lie on the side of (Y).

11.3. Summary

I have offered a general account of prioritarianism along the lines of accounts of egalitarianism and utilitarianism given elsewhere. These accounts make no use of quantitative evaluative measurements, and do not rely upon expected utility theory. I have given the most appealing defence of the priority view I can think of, and it has the bonus that fits with suggestive remarks made by Rabinowicz (2001, 2002) concerning the priority view and the separateness of persons. I cannot see a better defence. But ultimately, I have argued that it fails. Hence given (X), prioritarianism is prima facie implausible, and the defence offered on its behalf does not stand up to scrutiny. In most ways it is on a par with egalitarianism, which I claim is also prima facie implausible, and lacks a good defence (McCarthy, 2015). Given (Y), egalitarianism does not fail on

conceptual grounds. But I have argued that prioritarianism does. I have claimed that it may be indeterminate which of (X) and (Y) is correct, but that we ought to adopt (Y). Therefore I reject prioritarianism on conceptual grounds. But either way around, prioritarianism is not a viable theory.

Nevertheless, there should be some sort of explanation about why people have found it appealing. Aside from the possibility of a framing effect mentioned at the end of section 11.1, I suspect that many of us implicitly work with two welfare measures, something along the lines of the natural measure and the vNM measure. I also suspect that we are somewhat risk averse with respect to the natural measure. But if that is right, utilitarianism understood with respect to the vNM measure is consistent with prioritarianism understood with respect to the natural measure. The mistake was to take a natural idea out of context and to insert a reinterpretation of it into a complex theoretical framework where ordinary intuition is a poor guide.

ACKNOWLEDGEMENTS

This article was circulated and submitted in 2011, but for reasons I will pass over, I ceased work on it. I am very grateful to John Weymark for taking over the editorial process much later on, and for his patience, encouragement and good advice. The present version is more or less the same as the 2011 version, except that it borrows from a framework for thinking about utilitarianism developed in the meantime in joint work with Kalle Mikkola and Teruji Thomas. I thank an anonymous referee for two sets of helpful comments on the earlier version. Support was partially provided by the AHRC Research Leave Scheme and by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (HKU 750012H).

REFERENCES

Abdellaouia, M., C. Barrios, and P. Wakker. 2007. Reconciling introspective utility with revealed preference: experimental arguments based on prospect theory. *Journal of Econometrics* 138: 356–378.

Adler, M. 2011. *Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis*. New York, NY: Oxford University Press.

Atkinson, A. and J. Stiglitz. 1980. *Lectures on Public Economics*. London: McGraw-Hill.

Ben-Porath, E., I. Gilboa and D. Schmeidler. 1997. On the measurement of inequality under uncertainty. *Journal of Economic Theory* 75: 194–204.

Bouyssou, D. and J.-C. Vansnick. 1988. A note on the relationships between utility and value functions. In *Risk, Decision and Rationality*, ed. B. Munier, 103–114. Dordrecht: Reidel.

Broome, J. 1989. What's the good of equality? In *Current Issues in Microeconomics*, ed. J. Hey, 236–262. London: Macmillan.

Broome, J. 1990–91. Fairness. *Proceedings of the Aristotelian Society* 91: 87–102.

Broome, J. 1991. *Weighing Goods*. Oxford: Blackwell.

Broome, J. 2004. *Weighing Lives*. New York, NY: Oxford University Press.

Broome, J. 2015. Equality versus priority: a useful distinction. *Economics and Philosophy* 31: 219–228.
Brown, C. 2003. Giving up levelling down. *Economics and Philosophy* 19: 111–134.
Diamond, P. 1967. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility: comment. *Journal of Political Economy* 75: 765–766.
Fleurbaey, M. 2010. Assessing risky social situations. *Journal of Political Economy*, 118: 649–680.
Fleurbaey, M. 2015. Equality versus priority: how relevant is the distinction? *Economics and Philosophy* 31: 203–217.
Greaves, H. 2015. Antiprioritarianism. *Utilitas* 27: 1–42.
Hammond, P. 1991. Interpersonal comparisons of utility: why and how they are and should be made. In *Interpersonal Comparisons of Well-Being*, ed. J. Elster and J. Roemer, 200–254. Cambridge: Cambridge University Press.
Hardy, G., J. Littlewood and G. Pólya. 1934. *Inequalities*. Cambridge: Cambridge University Press.
Harsanyi, J. 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy* 63: 309–321.
Harsanyi, J. 1977. Nonlinear social welfare functions: a rejoinder to Professor Sen. In *Foundational Problems in the Special Sciences*, ed. R. Butts and J. Hintikka, 293–296. Dordrecht: Reidel.
Holtug, N. 2010. *Persons, Interests, and Justice*. New York, NY: Oxford University Press.
Hurley, S. 1989. *Natural Reasons*. Oxford: Oxford University Press.
Jackson, F. 2000. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. New York, NY: Oxford University Press.
Jensen, K. 1995. Measuring the size of a benefit and its moral weight. *Theoria* 61: 25–60.
Jensen, K. 2003. What is the difference between (moderate) egalitarianism and prioritarianism? *Economics and Philosophy* 19: 89–109.
Lewis, C. 1946. *The Analysis of Knowledge and Valuation*. La Salle, IL: Open Court Publishing.
Lewis, D. 1970. How to define theoretical terms. *Journal of Philosophy* 67: 427–446.
Lewis, D. 1973. *Counterfactuals*. Oxford: Blackwell.
McCarthy, D. 2006. Utilitarianism and prioritarianism I. *Economics and Philosophy* 22: 335–363.
McCarthy, D. 2007. Measuring life's goodness. *Philosophical Books* 48: 303–319.
McCarthy, D. 2008. Utilitarianism and prioritarianism II. *Economics and Philosophy* 24: 1–33.
McCarthy, D. 2013. Risk-free approaches to the priority view. *Erkenntnis* 78: 421–449.
McCarthy, D. 2015. Distributive equality. *Mind* 124: 1045–1109.
McCarthy, D., K. Mikkola and T. Thomas. 2016. Utilitarianism with and without expected utility. MPRA Paper No. 72578 https://mpra.ub.uni-muenchen.de/72578/.
McKerlie, D. 2001. Dimensions of equality. *Utilitas* 13: 263–288.
Meyer, M. and D. Mookherjee. 1987. Incentives, compensation, and social welfare. *Review of Economic Studies* 54: 209–226.
Myerson, R. 1981. Utilitarianism, egalitarianism, and the timing effect in social choice problems. *Econometrica* 49: 883–897.
Nagel, T. 1970. *The Possibility of Altruism*. Princeton, NJ: Princeton University Press.
Otsuka, M. and A. Voorhoeve. 2009. Why it matters that some are worse than others: an argument against the priority view. *Philosophy and Public Affairs* 37: 171–199.
Parfit, D. 1984. *Reasons and Persons*. Oxford: Clarendon Press.
Parfit, D. 1991. *Equality or priority? The Lindley Lecture*, University of Kansas.
Parfit, D. 2012. Another defense of the priority view. *Utilitas* 24: 399–440.
Pennings, J. and A. Smidts. 2000. Assessing the construct validity of risk attitude. *Management Science* 46: 1337–1348.
Rabinowicz, W. 2001. Prioritarianism and uncertainty: on the interpersonal addition theorem and the priority view. In *Exploring Practical Philosophy: From Action to Values*, ed.

D. Egonsson, J. Josefsson, B. Petersson and T. Rønnow-Rasmussen, 139–165. Aldershot: Ashgate.
Rabinowicz, W. 2002. Prioritarianism for prospects. *Utilitas* 14: 2–21.
Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
Samuelson, P. 1952. Probability, utility, and the independence axiom. *Econometrica* 20: 670–678.
Sen, A. 1976. Welfare inequalities and Rawlsian axiomatics. *Theory and Decision* 7: 243–62.
Sen, A. 1977. Non-linear social welfare functions: a reply to Professor Harsanyi. In *Foundational Problems in the Special Sciences*, ed. R. Butts and J. Hintikka, 297–302. Dordrecht: Reidel.
Suppes, P. and M. Winet. 1955. An axiomatization of utility based on the notion of utility differences. *Management Science* 1: 259–70.
Tungodden, B. 2003. The value of equality. *Economics and Philosophy* 19: 1–44.
Weymark, J. 1991. A reconsideration of the Harsanyi-Sen debate on utilitarianism. In *Interpersonal Comparisons of Well-Being*, ed. J. Elster and J. Roemer, 255–316. Cambridge: Cambridge University Press.

APPENDIX

In addition to the domain assumptions given in section 2, and the notation of section 2 and 9, we use the following terminology. For any $x \in \mathbb{W}$, 1_x is the prospect which assigns x a probability of one. For any $p \in \mathbb{P}$, $\mathcal{L}(p) \in \mathbb{L}$ is the lottery which (i) guarantees equality of outcome, and (ii) results in each $i \in \mathbb{I}$ facing the prospect p. For example, $\mathcal{L}(1_x)$ is the lottery which assigns the history $[x, \ldots, x]$ probability one, and $\mathcal{L}(\frac{1}{2}1_x + \frac{1}{2}1_y)$ is the lottery $[\frac{1}{2}, [x, \ldots, x]; \frac{1}{2}, [y, \ldots, y]]$.

Proof of Theorem 2. Assume (a). Define $\succsim_{PV}$ by $p \succsim_{PV} q \iff \mathcal{L}(p) \succsim \mathcal{L}(q)$. We first show $\succsim_{PV}$ is an outcome-equivalent strictly concave transform of $\succsim_{\mathbb{P}}$.

For monotonicity, using the Priority Principle then the definition of $\succsim_{PV}$:

$$1_x \succsim_{\mathbb{P}} 1_y \iff \mathcal{L}(1_x) \succsim \mathcal{L}(1_y) \iff 1_x \succsim_{PV} 1_y.$$

For strict concavity, suppose $1_x \succ_{\mathbb{P}} 1_y \succ_{\mathbb{P}} 1_z$. Then

$$1_y \sim_{\mathbb{P}} \tfrac{1}{2}1_x + \tfrac{1}{2}1_z \Rightarrow \mathcal{L}(1_y) \succ \mathcal{L}(\tfrac{1}{2}1_x + \tfrac{1}{2}1_z) \Rightarrow 1_y \succ_{PV} \tfrac{1}{2}1_x + \tfrac{1}{2}1_z.$$

By the definition of $\succsim_{PV}$, $\succsim$ satisfies Reduction to Prospects understood with respect to the preorder $\succsim_{PV}$. Since $\succsim$ also satisfies Anteriority and Two-Stage Anonymity, $L \succsim L' \iff p_L \succsim_{PV} p_{L'}$ follows from Theorem 1.

Conversely, assume (b). By Theorem 1, $\succsim$ satisfies Anteriority and Two-Stage Anonymity, and also Reduction to Prospects understood with respect to $\succsim_{PV}$. We wish to establish the Priority Principle.

For part (i) of the Priority Principle:

$$\mathcal{L}(1_x) \succsim \mathcal{L}(1_y) \iff 1_x \succsim_{PV} 1_y \iff 1_x \succsim_{\mathbb{P}} 1_y.$$

For part (ii), suppose $1_x \succ_{\mathbb{P}} 1_y \succ_{\mathbb{P}} 1_z$. Then

$$1_y \sim_{\mathbb{P}} \tfrac{1}{2}1_x + \tfrac{1}{2}1_z \Rightarrow 1_y \succ_{PV} \tfrac{1}{2}1_x + \tfrac{1}{2}1_z \Rightarrow \mathcal{L}(1_y) \succ \mathcal{L}(\tfrac{1}{2}1_x + \tfrac{1}{2}1_z).$$

■

Proof of Corollary 1. Assume (a). By Theorem 2, there exists an outcome-equivalent strictly concave transform $\succsim_{PV}$ of $\succsim_{\mathbb{P}}$ such that for all $L, L' \in \mathbb{L}$, $L \succsim L' \iff$

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press

$p_L \succsim_{PV} p_{L'}$. Since $\succsim$ satisfies EUT, it follows that $\succsim_{PV}$ also satisfies EUT. Hence there exists $f : \mathbb{W} \to \mathbb{R}$ such that

$$p \succsim_{PV} p' \iff E_p(f) \geq E_{p'}(f) \quad \text{for all } p, p' \in \mathbb{P}.$$

Since $\succsim_{PV}$ is outcome-equivalent to $\succsim_{\mathbb{P}}$, it follows that $f = w \circ u$ for some increasing function $w : I \to \mathbb{R}$.

Let $x, y \in I$ with $x > y$. Then there exist a, b and c in $\mathbb{W}$ such that $u(a) = x, u(b) = \frac{x+y}{2}$ and $u(c) = y$. These imply that $1_b \sim_{\mathbb{P}} \frac{1}{2}1_a + \frac{1}{2}1_c$ and $1_a \succ_{\mathbb{P}} 1_b \succ_{\mathbb{P}} 1_c$. Since $\succsim_{PV}$ is a strictly concave transform of $\succsim_{\mathbb{P}}$, these imply $1_b \succ_{PV} \frac{1}{2}1_a + \frac{1}{2}1_c$, and the displayed biconditional then implies $w(\frac{x+y}{2}) > \frac{1}{2}w(x) + \frac{1}{2}w(y)$. Since w is increasing, w is strictly concave on I (Hardy *et al*, 1934). But then, $L \succsim L' \iff p_L \succsim_{PV} p_{L'} \iff E_{p_L}(w \circ u) \geq E_{p_{L'}}(w \circ u) \iff E_L(W) \geq E_{L'}(W)$.

Conversely, assume (b). Define $\succsim_{PV}$ by $p \succsim_{PV} p' \iff E_p(w \circ u) \geq E_{p'}(w \circ u)$. Then we have $L \succsim L' \iff E_L(W) \geq E_{L'}(W) \iff E_{p_L}(w \circ u) \geq E_{p_{L'}}(w \circ u) \iff p_L \succsim_{PV} p_{L'}$.

We now need to show $\succsim_{PV}$ is an outcome-equivalent strictly concave transform of $\succsim_{\mathbb{P}}$. For outcome-equivalence: $1_x \succsim_{\mathbb{P}} 1_y \iff u(x) \geq u(y) \iff w(u(x)) \geq w(u(y)) \iff E_{1_x}(w \circ u) \geq E_{1_y}(w \circ u) \iff 1_x \succsim_{PV} 1_y$.

For strict concavity, suppose $1_x \succ_{\mathbb{P}} 1_y \succ_{\mathbb{P}} 1_z$. Then $1_y \sim_{\mathbb{P}} \frac{1}{2}1_x + \frac{1}{2}1_z \Rightarrow u(y) = \frac{1}{2}u(x) + \frac{1}{2}u(z) \Rightarrow w(u(y)) > \frac{1}{2}w(u(x)) + \frac{1}{2}w(u(z)) \Rightarrow E_{1_y}(w \circ u) > E_{\frac{1}{2}1_x + \frac{1}{2}1_z}(w \circ u) \Rightarrow 1_y \succ_{PV} \frac{1}{2}1_x + \frac{1}{2}1_z$.

Hence $L \succsim L' \iff p_L \succsim_{PV} p_{L'}$, and $\succsim_{PV}$ is an outcome-equivalent strictly concave transform of $\succsim_{\mathbb{P}}$. Generalized prioritarianism follows from Theorem 2. ■

BIOGRAPHICAL INFORMATION

David McCarthy teaches at the University of Hong Kong, and mainly works on ethics, decision theory and formal epistemology.

https://doi.org/10.1017/S0266267116000225 Published online by Cambridge University Press